Molecular bases of equine polysaccharide storage myopathies

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Raffaella Bertoni Cavalcanti Teixeira

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Molly McCue, James Mickelson

April, 2015

Acknowledgements

First of all I am thankful to Drs McCue and Mickelson for their essential support. Without their superior knowledge and experience I would not have been able to accomplish this work.

The author would also like to thank Drs Reed, Rendahl and Valberg for being part of my committee and for their support and help.

I would also like to acknowledge Rob Schaefer for all his help with scripts and commands. I have learned a lot from him. I would like to thank my lab mates Sam Beeson, Elaine Norton, Annette McCoy, Felipe Avila, Nichol Schultz and Ann Kemper for their help and friendship.

And last, I want to thank my dog Stella for showing me her love during good and bad times.

I have met incredible people during this journey and I'm very thankful to each one of them. You have all left a mark on my life forever.

Dedication

This thesis is dedicated to my adviser Dr McCue for helping me achieve my goal of completing a PhD. Dr McCue has taught me so many important lessons. I'm a better scientist, clinician and most of all, a better person because of her. Her lessons will stay with me for the rest of my life.

I also dedicate this work to my co-adviser, Dr Mickelson, for being supportive, helpful, wise and very patient.

I dedicate this thesis to my family (Mom, Dad, brother and sister-in-law). You are the most important thing I have in my life. Your unconditional support gives me strength to fight for my dreams.

And last, I would like to dedicate this thesis to my friends, especially Fernanda Shoyama, Marina Figueiredo and Alex Draper for always being there for me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Literature Review

Polysaccharide Storage Myopathy

Polysaccharide Storage Myopathy (PSSM) is a form of glycogen storage disease in horses, characterized by abnormal polysaccharide inclusions in skeletal muscle[1]. Clinical signs consistent with PSSM were first described in working draft horses after a period of rest in the 1900s [1-3]. It was believed that clinical signs were a consequence of increased lactate accumulation in the muscle secondary to excessive accumulation of muscle glycogen during periods of rest. In 1992, PSSM was first described as a muscular glycogen storage disorder in Quarter Horses (QH) and related breeds[2]. The disease has now been described in many different breeds and countries. PSSM is a common diagnosed equine muscle disorder and can lead to poor performance, loss of athletic function and can be potentially life threatening[2,4-8].

Clinical Signs

The severity of clinical signs can vary from decreased performance to severe signs of acute rhabdomyolysis. Quarter Horses and related breeds commonly develop muscle stiffness and pain, shifting lameness, sweating, reluctance to move, colic-like signs, camped-out stance, and in some cases myoglobinuria and recumbency[1]. In draft horses common clinical signs include muscle atrophy and progressive weakness, while gait abnormalities are commonly seen in Warmbloods [8,9].

Signs are usually first noted at 2-3 years of age, when training begins. Subsequently, the disease is manifested when exercise is implemented after a period of rest. PSSM adult horses or young foals that are exposed to infectious diseases may have exacerbation of the signs and can develop rhabdomyolysis[10]. Post-anesthetic complications have also been reported[11].

Muscle Histopathology

Interpretation of the amount and type of polysaccharide in muscle biopsy samples can be subjective and two different laboratories initially used two different criteria. The less restricted approach includes horses with amylase-*sensitive* glycogen and sarcoplasmic masses[6]. Since including horses with amylase-sensitive polysaccharide

increases sensitivity of the test but markedly decreases specificity, the suggested gold standard diagnosis of PSSM in muscle biopsy was the presence of periodic acid Schiff (PAS) positive inclusions of amylase-*resistant* polysaccharide in type 2A and type 2 B muscle fibers [2,12]. Other histopathologic findings include centrofascicular atrophy and necrosis of type 2A and 2B muscle fibers. Electron microscopy shows increased filamentous material, large accumulation of beta glycogen between myofibrils and under the sarcolemma, mild to severe myofibrillar lysis and degenerate mitochondria containing electron dense inclusions [13,2] False negative PSSM cases can be found in young animals based on absence of amylase-resistant polysaccharide in muscle histopathology [14], since it can take up to 2 years for abnormal glycogen to start to accumulate. Although cardiac abnormalities are not a common feature in PSSM horses, evidence of abnormal polysaccharide has been identified in a few severely affected horses[15-17].

Prevalence

Based on histopathology, PSSM has been described in many different breeds and is the most common muscular disorder in Draft Horses, QH and related breeds, and Warmbloods[8,18,19]. The overall prevalence of PSSM in horses with muscular disorders severe enough to prompt a muscle biopsy is 21.7% when only horses with PAS positive amylase-*resistant* polysaccharide inclusions in the muscle are considered. When a less stringent diagnostic criteria is used, and horses with abnormal amylase-*sensitive* polysaccharide in muscle fibers are considered, the prevalence increases from 21.7% to 40.1%. The prevalence of PSSM among healthy QHs based on skeletal muscle biopsy was estimated to be between 6% and 12%[20]. Several QHs with abnormal polysaccharide lacked clinical signs of PSSM, likely due to management practices[20], which in turn could lead to increase prevalence of the disease in the population by breeding affected asymptomatic horses. In North American Belgian Horses, a prevalence of 34% has been suggested based on amylase-resistant polysaccharide in muscle biopsies of 113 horses[18], while in the UK the prevalence of PSSM in Draft Horses was reported to be 8% [7]. Different from recurrent exertional rhabdomyolysis, no gender predisposition has been described in horses with PSSM[1].

Biochemistry and metabolic findings related to the energy deficit

In humans, glycogen storage myopathies are commonly associated with inability to metabolize glycogen or its metabolites (glycogenolysis and glycolysis). This possibility was investigated in horses and no enzymatic defect similar to any glycogen disorder described in humans was found[2,21,22]. PSSM horses have 1.5 to 2 times normal glycogen concentrations in the muscle compared with controls, but have the ability to breakdown glycogen during exercise and produce lactate[21]. Muscle energy metabolism is closely regulated by the cytosolic ATP:ADP ratio. A decrease in this ratio stimulates oxidative metabolism, followed by myokinase reaction if ATP can't be restored. The myokinase reaction combines 2 ADP and produces ATP and adenine monophosphate (AMP), which will activate AMP deaminase and consequently increase levels of muscle inosine monophosphate (IMP). Increased levels of IMP in the muscle usually occur during intense exercise or under metabolic stress. PSSM horses, during submaximal exercise (which relys mostly on aerobic metabolism), show disruption of cellular energy metabolism and considerable metabolic stress, characterized by increased IMP levels in single muscle fibers and increased serum CK activity[23]. This IMP elevation after submaximal (~ 20 minute) exercise is comparable with horses performing at maximal exercise. Even with as little as 11 minutes of walk and trot, horses with PSSM have adenine nucleotide degradation in some fibers that resembles horses performing at twice the speed for almost 5 times as long. During light submaximal exercise, healthy horses should not have any adenine depletion[23,21]. The mechanism for this energy deficit remains unknown.

Discovery of the *GYS1* mutation – Type 1 PSSM

Candidate gene studies of various types appeared to exclude an association of PSSM with the muscle isoform of phosphofructokinase (*PKFM*); the AMP activated protein kinase (*AMPK*); the glucan (1,4-alpha); branching enzyme 1 gene - *GBE1*

(encodes a glycogen branching enzyme); amylo-alpha-1, 6-glucosidase; 4-alpha-glucanotransferase gene – *AGL* (encodes a glycogen-debranching enzyme), [24].

Discovery of the GYS1 mutation

Genome-wide association analysis was performed in a population of 48 related PSSM and 48 unrelated Quarter Horses controls. Related cases were chosen to maximize the level of linkage disequilibrium (they did not share more genetic material than second cousins, to avoid false inflation of association due to sampling of closely related individuals). PSSM cases were selected based on muscle histopathology findings consistent with the presence of PAS-positive, amylase-resistant inclusions of polysaccharide in gluteal or semimembranous muscle fibers and the presence of excessive glycogen levels in the muscle and/or a clinical history of exertional rhabdomyolysis. Controls were selected based on normal muscle biopsy, normal muscle glycogen concentrations and no history of exertional rhabdomyolysis. Controls were not related to the case cohort and were not more related than third cousins[25].

One hundred and five microsatellite (MS) markers, distributed on all 31 equine autosomes, were used. Two markers, in two different chromosomes, had significant difference in allele distribution between cases and controls, based on Pearson's X^2 tests of independence. The marker on chromosome 10 had the lowest p-value and its allele distribution showed that 23 out of 48 affected horses had an allele virtually absent in controls. Eight additional ECA10 MS were genotyped and defined a 10-Mb region highly associated with PSSM. To narrow this candidate region and to account for possible association due to relatedness between cases and controls, an additional 52 PSSM cases (determined not to share common ancestor with initial case-cohort within 5 generations) and 44 controls were genotyped. Two markers within this follow up population were significantly associated with PSSM in the combined case and control populations and represented a 3-Mb region in chromosome 10. This region is homologous to human chromosome 19, where the only gene related to carbohydrate metabolism is the glycogen synthase 1 gene (*GYS1*). Genomic DNA or cDNA were isolated from a PSSM and a control horse and the *GYS1* gene was amplified and sequenced. A single nucleotide

polymorphism was identified in the PSSM horse (G to A base pair substitution that changes the amino acid arginine to histidine on codon 309 – Arg309His). To further determine if this was the PSSM causative mutation, 6 additional cases and 6 controls were genotyped for the Arg309His mutation by restriction fragment length polymorphism assay. Briefly, the *GYS1* exon 6 and the flanking intronic region are amplified and a restriction enzyme cuts the *GYS1* H allele at the site of the mutation and at an intronic site present in both the R (wild-type) and A (mutant) alleles. None of the controls had the Arg309His mutation, while 4 cases were heterozygous and 2 were homozygous for the mutation. The non-mutated arginine 309 is highly conserved in the *GYS1* and *GYS2* forms of glycogen synthase in several different species, indicating that the region has an important function[25].

An additional 99 PSSM QHs, used in the GWAS analysis were genotyped for the *GYS1* mutation. While most of the horses were heterozygous (n=72) or homozygous (n=5) for the mutation, 22 PSSM horses were homozygous for the wild-type allele R, suggesting that a second form of PSSM (non-*GYS1*) was present or that the *GYS1* was not the causative mutation. Sequencing from 480 bp upstream of the predicted promoter region through exon 1, the entire coding sequence and the 3'UTR showed no additional sequence difference between cases and controls. Also, association analysis using the same ECA10 MS markers excluded this chromosomal locus and *GYS1* as the cause of PSSM in QH homozygous for the *GYS1* wild-type allele. The vast majority of control horses were homozygous for the R allele, although a few horses were heterozygous for the H allele, which might be due to false negatives based on muscle histopathology[25], or due to incomplete penetrance of the H allele.

The *GYS1* gene encodes the muscle glycogen synthase isoform; the limiting step in glycogen synthesis. The Arg309His mutation results in enhanced activity and/or poor regulation of the enzyme, leading to increased muscle glycogen concentrations. PSSM horses have 1.5 to 2 times more muscle glycogen compared to healthy controls[26]. The activity of the glycogen synthase enzyme is not highly regulated by *GYS1* expression in mammals. Protein kinases (by phosphorylation), protein phosphatase-1 and allosteric stimulation by glucose-6-phosphate are important regulators of the enzyme activity[27] . Glycogen synthase assays on muscle homogenates revealed increased activity of the

enzyme in PSSM comparing with control muscle, in the presence or absence of the allosteric activator glucose 6-phosphate (G6P) [25].

Haplotype analysis in several different breeds that carry the *GYS1* mutation, revealed that all chromosomes with the mutated *GYS1* A allele had a single 351kb conserved haplotype, suggesting that the A allele is identical by descent in all horse breeds and the mutation arose in a single horse before the formation of the modern horse breeds.

In summary, the *GYS1* mutation causes type 1 polysaccharide storage myopathy and is a dominant gain-of-function mutation present in the coding sequence of exon 6, which results in poor regulation and increased activity of the skeletal muscle glycogen synthase isoform, the limiting step on glycogen synthesis. The mutation may result in the enzyme being in a continuously active state, which in turn could be responsible for the accumulation of excessive glycogen and abnormal amylase-resistant polysaccharide in horses with type 1 PSSM [28].

Variable phenotypic expression in type 1 PSSM horses

Environmental factors largely influence clinical phenotype of horses with PSSM. Controlled regular exercise, increased turn-out time, and a diet low in non-structural carbohydrates and high in fat concentration, have been successful in decreasing muscle stiffness in all PSSM horses and eliminating episodes of rhabdomyolysis in 75% of the cases[1,29,30].

The amount and type of fat that should be provided to PSSM horses have been evaluated. Providing fat with an odd number of carbon atoms (triheptanoin; short-chain fatty acid) has been used in humans with metabolic myopathies, to restore citric acid cycle intermediates (CAC) by providing a different source of energy (succinyl-COA)[30]. Surprisingly, in PSSM horses, this diet caused detrimental effects by possible stimulation of insulin secretion (that leads to increase glycogen synthesis and inhibition of lipolysis suggested by decreased availability of NEFA). Feeding long-chain fatty acids (corn oil) significantly decreased exercise intolerance and rhabdomyolysis in PSSM horses, by

providing energy for aerobic exercise[108 30]. This energy was provided by increasing the availability of fat for skeletal muscle oxidation, and possibly increasing glycogenolytic, glycolytic and oxidative flux[30].

Even when those environmental factors are controlled, phenotypic variation in the severity of the disease is observed. Evidence that PSSM1 is an incomplete penetrant dominant trait has been shown in Draft horses[15], Quarter Horses[25] and Warmbloods[31]. Not all horses heterozygous for the *GYS1* mutation show clinical signs of the disease.

An additive effect of the mutated allele was demonstrated in Belgian and Percheron horses, where type 1 PSSM cases with 2 copies of the mutated allele had increased amounts of amylase-resistant polysaccharide in type 2A muscle fibers compared with heterozygotes, followed by control horses (homozygous normal allele). The changes in muscle histopathology were positively correlated with increased activity of resting serum aspartate transaminase (AST). Homozygous draft horses had higher serum muscle enzyme activities (CK and AST) at rest, followed by heterozygotes and controls, but post exercise, serum muscle enzymes were not depend on allele copy number. Post-exercise, homozygous draft horses had higher creatine kinase (CK) activity than heterozygotes and controls, but no significant difference in serum CK activity was observed between heterozygotes and control horses. AST was not significantly different after exercise between homozygotes, heterozygotes and controls. The severity of muscle histopathology appears to depend on underlying genotype [15].

The presence of modifying genes can also cause phenotypic variation. An identified family of Quarter Horses with a severe form of PSSM, characterized by signs of recurrent episodes of rhabdomyolysis and sudden death, was less likely to respond to recommended changes to diet and training[32]. The severity of the signs in this family could not be explained by homozygosity for the *GYS1* mutation, since most of the horses had just one copy of the mutated allele. The presence of a mutation in the ryanodine receptor 1 - skeletal muscle calcium release channel (*RYR1*) gene in this family, which causes malignant hyperthermia, in conjunction with the *GYS1* mutation, led to a more severe PSSM phenotype[32]. This is likely due to more marked muscle damage secondary to excessive calcium release from the sarcoplamsic reticulum in muscle cells (caused by

the *RYR1* mutation) in conjunction with muscle energy deficit during exercise in horses with the *GYS1* mutation [32].

Prevalence of the GYS1 mutation

In a study of 750 horses diagnosed with PSSM based on muscle biopsy, 350 horses from 15 different breeds were either homozygous or heterozygous for the *GYS1* mutation. Therefore, 394 PSSM horses were not carriers of the *GYS1* mutation[25]. In a second study of 901 horses from 36 different breeds diagnosed with PSSM by muscle biopsy, the *GYS1* mutation was found in 50% of the cases and in 17 different breeds [9]. The mutated H allele had a frequency of 0.35, the wild-type R allele had a frequency of 0.65 and the majority of the horses were heterozygous for the mutation[9].

The highest prevalence was found in Belgian and Percheron horses (80% of PSSM cases that had amylase resistant polysaccharide in the muscle biopsy), with an 11% incidence of homozygous draft horses. In QH and related breeds, the mutation is responsible for approximately half of PSSM cases diagnosed by muscle biopsy[9]. Warmblood horses with PSSM based on muscle biopsy had a lower prevalence of the *GYS1* mutation at 18% and none of the Thoroughbred, Arabians and Standardbred horses diagnosed with PSSM by muscle biopsy were heterozygous or homozygous for the mutation[9]. In a study in the UK, 65% of horses with PSSM diagnosed by muscle biopsy were heterozygous for the *GYS1* mutation[33].

So far, the mutation has been found in more than 20 breeds of horses from North America and Europe[9,33], including Quarter Horses, Paint Horses, Appaloosa and Draft Horses (Belgian, Haflinger, Percheron, Shire, Suffolk Punch, American Cream Draft), Warmbloods (Hanoverian), Morgan, Mustang, Tennessee Walking horse, Rocky Mountain horse breeds, Cob Normand draught horses, mixed breeds and others [9,25,34]. The frequency of the mutant allele was the highest in the North American Percheron and Belgian Draft breeds followed by the South German Coldbloods, QHs and related breeds. The mutation was found in low frequencies in Shires and Morgans[35]. The presence of the *GYS1* mutation was also evaluated in 403 horses from 13 different European draft horse breeds from Belgium, France, Spain, Germany, Sweden and the Netherlands. While the true prevalence of the disease could not be assessed in the study since the horses were not

randomly selected, the mutation was found in all 13 breeds (Ardenner, Belgian draft horse, Breton horse, Comtois, Trait du Nord, Hispano-Breton, Netherlands draft horse and German coldbloods) and all 6 countries. Sixty two per cent of the horses (250 of 403 horses) were heterozygous for the *GYS1* mutation and several horses within 8 of the 13 breeds were homozygous for the mutation [36].

A family of Warmblood horses in Switzerland, with a high prevalence of exertional rhabdomyolysis (ER), was evaluated for the presence of abnormal polysaccharide accumulation on muscle histopathology and the presence of the *GYS1* mutation. The sire and 71 of his descendants were investigated. Horses were in training for at least 1 year. Signs of ER were observed in 39% of the horses and the *GYS1* mutation was present in 51% of the 71 descendants, indicating that not all horses that are heterozygous for the type 1 PSSM genotype manifest signs of the disease. In this family, 41% (15/37) of *GYS1* positive horses did not have signs of ER. Risk factors associated with the manifestation of clinical signs include the presence of the *GYS1* mutation (7.1-times increased odds ratio for being affected by ER in comparison with wild-type horses) and gender. In contrast with previous studies[1,19,33], females were more likely to develop signs of ER in the horses with a history of ER and horses with the *GYS1* mutation. One possible explanation for the increased risk observed in *GYS1* positive mares, is the possible presence of a concomitant muscle disorder, such as RER. Age and management factors (turn out time, frequency of use an amount of concentrate feed) had no effect on ER development in the *GYS1* positive horses, but the specific amount of starch in the concentrate feed could not be evaluated [37].

These findings confirm that the *GYS1* mutation is not the only cause of PSSM and a second, non-*GYS1* glycogenosis is present. PSSM horses that are not heterozygous for the *GYS1* mutation are referred to as having ***type 2 PSSM.***

With the association of PSSM with clinical disease, the large number of breeds that possess the *GYS1* mutation and the high prevalence of the mutation in some breeds is surprising. However, previous studies have demonstrated that clinical signs of PSSM can be ameliorated with consistent daily exercise and limited starch intake[29], conditions that were typical for horses from the time of domestication until their decline as a utilitarian animal after World War II. Under these historical management conditions, the ability to

rapidly restore muscle glycogen after exercise without supplemental feeding would create a more efficient work animal. Therefore, McCoy *et al* [38] hypothesized that the *GYS1* mutation represents a thrifty genotype that was driven to high frequency by artificial selection[25].

McCoy and colleagues investigated a 1.97 Mb region around the *GYS1* mutation for evidence of positive selection in Quarter Horses and Belgian horses. Seventy-two horses from 5 different domestic breeds, 1 domestic ass and 2 Przewalski's horses were Sanger sequenced around the *GYS1* mutation (1606bp upstream and 1751bp downstream; 3 homozygous horses, 49 heterozygous and 26 wild-type) and 9 novel intronic SNPs were identified. Phasing of this data revealed that all chromosomes that had the mutated A allele were found on a single haplotype across the 9 novel SNPs and the *GYS1* mutation. Chromosomes containing the wild-type G allele occurred as 8 distinguishible haplotypes. In addition 279 horses, from 8 breeds (including 51 horses from the initial Sanger sequencing) were genotyped for 51 SNPs in a 1.97–Mb region around the *GYS1* mutation. The markers include the *GYS1* mutation itself, 41 SNPs previously discovered in the region by Sanger sequencing at the Broad Institute and 5 of the 9 SNPs discovered in the original 72 horses. Samples included 166 PSSM1 cases and 113 controls (179 chromosomes with the A allele and 379 with the G allele). The genotypes were phased, imputation was performed for missing data and haplotype frequency was estimated. A highly conserved haplotype of 350Kb and 20 SNPs surrounding the *GYS1* mutation was observed in chromosomes containing the mutated A allele (80% of the chromosomes containing the A allele had a single haplotype). The chromosomes containing the G allele contained 9 haplotypes. Relative extended haplotype homozygosity (REHH) for the A haplotype was increased for Belgians horses in comparison with the G haplotypes even when comparing with simulated data under neutral drift, suggesting that the high frequency of the mutated allele in this population is unlikely due to a bottleneck process alone. Identifying genomic regions with unexpectedly high local haplotype homozygosity relative to neutral expectation represents a powerful strategy to ascertain candidate genes responding to natural or artificial selection. EHH is defined as the probability that two randomly chosen chromosomes carrying the same allele at a focal SNP are identical by descent over a given distance surrounding it. Different chromosomal regions can have

different recombination rates, which affects linkage disequilibrium (LD) (LD is stronger in regions where recombination is not common than in regions that exhibit elevated rates of recombination). So a greater measure of LD could originate from low recombination rates in a particular region, rather than recent positive selection. To take this issue into consideration, REHH measures LD surrounding a haplotype of interest to minimize the regional recombination effects. Excess muscle glycogen might have been advantageous under conditions of scare feed and strenuous exercise, which differs dramatically from modern management conditions that lead to precipitation of clinical signs. [39].

Hypothesized link between GYS1, glycogen accumulation and rhabdomyolysis

PSSM horses, during submaximal aerobic exercise, show disruption of cellular energy metabolism and considerable metabolic stress, characterized by increased IMP levels in single muscle fibers[23]. While a gain of function in glycogen synthase (GS) explains the muscle biopsy phenotype in type 1 PSSM, the link between excessive muscle glycogen, abnormal polysaccharide and rhabdomyolysis during sub-maximal exercise is less clear. PSSM1 horses can metabolize glycogen and have a normal flux of metabolites through glycolysis during maximal exercise[21], yet these horses demonstrate exercise intolerance, painful muscle cramping, and rhabdomyolysis during sub-maximal (aerobic) exercise.

Carbohydrates and fatty acids are the two major donors for aerobic phosphorylation. During the first 30 minutes of submaximal exercise most of the energy is provided by muscle glycogen[40]. Circulating FFA concentrations increase within 15 minutes of low-intensity exercise and uptake of short-chain and medium chain FFA across muscle membranes occurs down a concentration gradient by diffusion. Long-chain fats are translocated across the muscle membrane by active transport (fatty acid transporters). Transport of FFA from the cytoplasm into mitochondria for beta-oxidation is facilitated by an acylcarnitine transport system. A continuous supply of pyruvate from glucose or glycogen is essential for metabolism of FFA. Glycogen provides pyruvate which replenishes intermediates in the citric acid cycle allowing acetyl coenzyme A (CoA) derived from beta-oxidation of FFA to enter the first step of the citric acid cycle[40].

The energy deficit observed during submaximal exercise in PSSM horses might be secondary to the constant activation of the GS enzyme (due to the gain of function mutation in *GYS1* gene and exacerbation by insulin release when consuming grain), which might be interpreted by the cell as an indication that glycogen breakdown and lipolysis are not necessary. Normally, the cell would sense an energy deficit and AMP kinase activation would stimulate oxidative metabolism of fatty acid and carbohydrate by activation of pyruvate dehydrogenase. If pyruvate dehydrogenase is not fully activated adequate acetyl CoA for oxidative metabolism is not produced. If PSSM horses are on a grain diet, they have low plasma FFA concentrations[40], possibly due to suppression of lipolysis by high insulin. So fatty acid oxidation, also fails in supplying acetyl CoA. Also, PSSM horses on a grain diet have high muscle citrate concentrations which might inhibit generatation of sufficient acetyl CoA for muscle energy from carbohydrate or fat metabolism during submaximal exercise [108 30]. High citrate concentrations activate acetyl CoA carboxylase that converts acetyl CoA to malonyl CoA which directs acetyl CoA away from the citric acid cycle and causes inhibition of carnitine palmytoyl transferase (CPT1), the key enzyme necessary to transport long-chain fatty acids into the mitochondria for β-oxidation[40].

Current recommendations to control rhabdomyolysis in PSSM horses include regular exercise, increased turnout time, and a diet with low non-structural carbohydrates and high fat concentration. The provision of long-chain fatty acids in the diet of PSSM horses significantly decreases exercise intolerance and rhabdomyolysis with exercise[1 29]. Fat supplementation increases plasma FFA concentrations and the availability of fat for oxidation in skeletal muscle. It has been shown that the contribution of FFAs to energy production during low-intensity exercise is largely dependent on supply[40]. Daily exercise may lower plasma insulin and increase plasma FFA concentrations, and training over time may enhance uptake of fatty acids into skeletal muscle and improve muscle oxidative capacity[40].

Research questions to be addressed for PSSM1

This section of the thesis will address the energy deficit responsible for rhabdomyolysis in type 1 PSSM horses.

Hypothesis and Specific Aims: PSSM1

Skeletal muscle energy deficit in type 1 PSSM

Rationale: While a gain of function in glycogen synthase (GS) explains the muscle biopsy phenotype in type 1 PSSM, the link between GS, increased muscle glycogen content and clinical signs is unknown. PSSM1 horses can metabolize glycogen and have a normal flux of metabolites through glycolysis during maximal exercise[21], yet these horses demonstrate exercise intolerance, painful muscle cramping, and rhabdomyolysis during sub-maximal (aerobic) exercise [12,17,21,23], demonstrated by high IMP concentrations secondary to a decrease in the cytoplasmic ATP/ADP ratio and triggering of the myokinase reaction in an attempt to restore cellular ATP levels[41]. There are a number of cellular mechanisms that couple cellular energy balance and flux of substrates into aerobic metabolism with the intensity, duration and frequency of exercise[42]. One of the critical mechanisms may be cellular glycogen concentration and/or glycogen synthase activity.

Central Hypothesis: Skeletal muscle energy deficit and rhabdomyolysis in type 1 PSSM horses is due to the effects of altered regulation of gene expression in energy producing metabolic pathways induced by excessive glycogen levels and\or improperly regulated GS activity.

Secondary hypothesis: Rhabdomyolysis in PSSM1 horses can be managed in large part by consistent controlled daily exercise[1,29,43], which increases the muscle's capacity to oxidize fat [23,29,44]. Extensive research in other species has demonstrated that exercise triggers nutrient sensing molecules such as AMP kinase, which in turn leads to altered

expression of several downstream targets[42]. This leads to the secondary hypothesis that the clinical improvement in PSSM1 horses resulting from consistent daily exercise is due to restoration of expression patterns of genes in major metabolic regulatory pathways.

Aim 1: Investigate the genes and gene pathways that are altered in PSSM1 horses during subclinical rhabdomyolysis, and the altered expression patterns after training by quantifying gene expression using next generation sequencing of the skeletal muscle transcriptome (RNA-seq) in PSSM1 cases and controls.

RNA-seq

Whole genome next generation RNA sequencing is a high throughput sequencing technology of cDNA. It has several advantages over technologies for transcriptome analysis, such as microarray. It achieves base-pair level resolution, a higher dynamic range of expression levels, it does not rely on probe design and previous genome knowledge and therefore is not limited to detecting known transcripts and isoforms and permits de novo transcriptome assembly when a high quality reference genome or no reference genome is available. A few concerns have to be taken into consideration when performing RNA-seq, including how uniformly sequences from an entire transcript will be represented, how much sequence is required to detect and measure the transcripts with low abundance and appropriate downstream analysis since expression levels rely on proper mapping of sequencing reads to corresponding reference genome or on their efficient de novo assembly [45-48]. Pathway analysis is then performed to gain insights into the underlying biology of differentially expressed genes at the function level. Common diseases often arise from the joint action of multiple SNPs/genes within a pathway. Although each single SNP may confer only a small disease risk, their joint actions are likely to have a significant role in the development of disease[49].

Type 2 PSSM

Clinical signs

Both, type 1 and type 2 PSSM horses have similar clinical presentation, with no significant difference in the incidence of exertional rhabdomyolysis, muscle fasciculation or muscle atrophy[12]. Horses with type 2 PSSM tend to present more often with an undiagnosed gait abnormality compared to *GYS1* positive horses [12].

Prevalence

In a study of QHs and related breeds, type 2 PSSM cases tend to be younger than type 1, with 32% being younger than 1 year old (against 6% of PSSM1 horses)[12]. 27% of QH diagnosed with PSSM based on muscle biopsy do not have the *GYS1* mutation[9] and a large proportion of Warmbloods with PSSM are not carriers of the *GYS1* mutated allele[9,12].

Biopsy findings

Muscle biopsy findings in type 1 and 2 PSSM horses differ mainly in the appearance and location of abnormal polysaccharide. Abnormal polysaccharide in type 1 PSSM horses is often amylase resistant, course granular in appearance and commonly located in the cytoplasm. In type 2 PSSM horses the abnormal polysaccharide is often amylase-sensitive, fine granular and homogenous in appearance and commonly located just below the sarcolemma[12]. No significant difference in muscle atrophy, myonecrosis, central located nuclei and the amount of abnormal polysaccharide was observed between type 1 and 2 horses. On electron microscopy, *GYS1* positive and negative PSSM horses contained large amounts of β glycogen particles in the cytoplasm and subsarcolemmal. The abnormal granular amylase-resistant polysaccharide in *GYS1* positive horses had distinctive polyglucosan bodies, consisting of filamentous material surrounded by β glycogen particles. Immunohistochemistry showed strong staining for desmin in regions containing the granular abnormal polysaccharide in type 1 PSSM horses, while minimal staining was observed in controls and type 2 PSSM horses. Myoglobin was co-localized

with course granular polysaccharide in type 1 PSSM horses. Polyglucosan is a less highly branched β glycogen structure and contains several protein structures including myoglobin, desmin and ubiquitin [50] [12]. The abnormal β glycogen particles observed in type 2 PSSM horses were not membrane bound, and in association with the absence of polyglucosan, is consistent with other reported glycogenoses in humans where the ratio of branched-to-unbranched glycogen is not affected[12].

Preliminary GWAS analysis

Genome wide association analysis (GWAS) involves the study of several genetic markers in the genome and their association with a trait. The goal of population association studies is to identify polymorphisms that vary systematically between cases and controls. In GWAS analysis no assumptions are made and association analyses are performed in the entire genome. GWAS analysis relies on the concept of LD, which is the non-random association between markers and the causal allele. There is a low chance of recombination between two loci that are close together within a chromosome segment, making it possible to find markers that are associated with disease in case-control GWAS analysis. In this type of study, population structure can generate false positive genotype-phenotype association and should be accounted for during the analysis, using mixed model approaches. Correction for multiple testing should also be performed, to decrease the chance of rejecting the null hypothesis (no association between marker and phenotype) by chance[51-54].

To identify regions of the genome associated with the PSSM2 phenotype, a cohort of 228 horses (104 PSSM2 cases and 124 controls) was genotyped for 54,602 SNP markers with the Illumina Equine SNP50 Beadchip and genome wide association analysis was performed. GWAS consisted of both basic association tests and structured association analysis to account for population structure within the sample cohort. Basic (unstructured) association included allelic association (chi-square) and additive, genotypic, recessive and dominant logistic regression models. Dominant logistic regression identified six SNPs on equine chromosome 18 (ECA18) ($p=9.44 \times 10^{-7}$ to 2.84×10^{-6}) and a single SNP on ECA3 (2.99×10^{-6}) that were associated with PSSM2.

Association with these SNPs persisted after correction for multiple testing using 10,000 label-swapping permutations (Figure 1). To account for population structure present in the sample cohort (data not shown), structured associations were performed using mixed model, and combined mixed model and principal components analysis. Logistic regression with 3 principal components and a mixed model with a relationship matrix derived from whole genome SNP data confirmed initial GWAS results for the ECA18 SNPs (1.29 x10-6 to 7.41 x 10-6) (Figure 2). The ECA3 SNP was not genome-wide significant after correction for population structure.

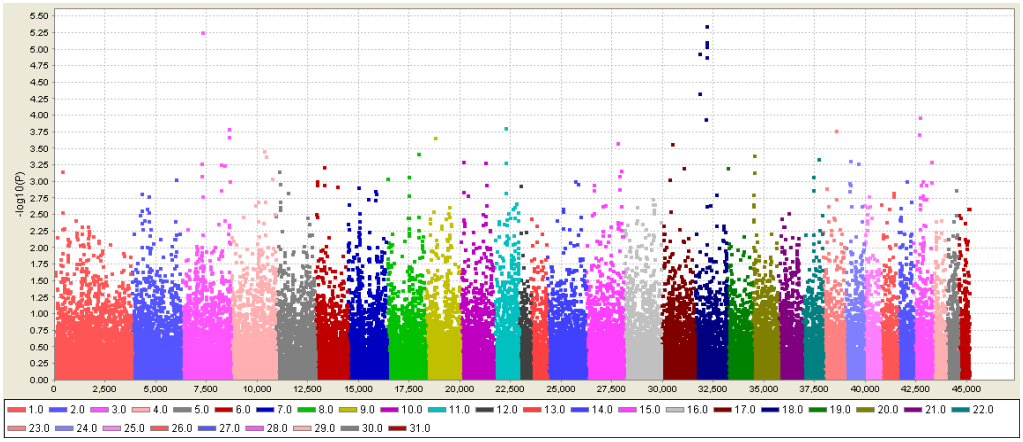


Figure 1: **Manhattan plot PSSM2 GWAS.** Chromosomes are plotted across the horizontal axis, and –log p-value for individual SNPs plotted on the vertical axis.

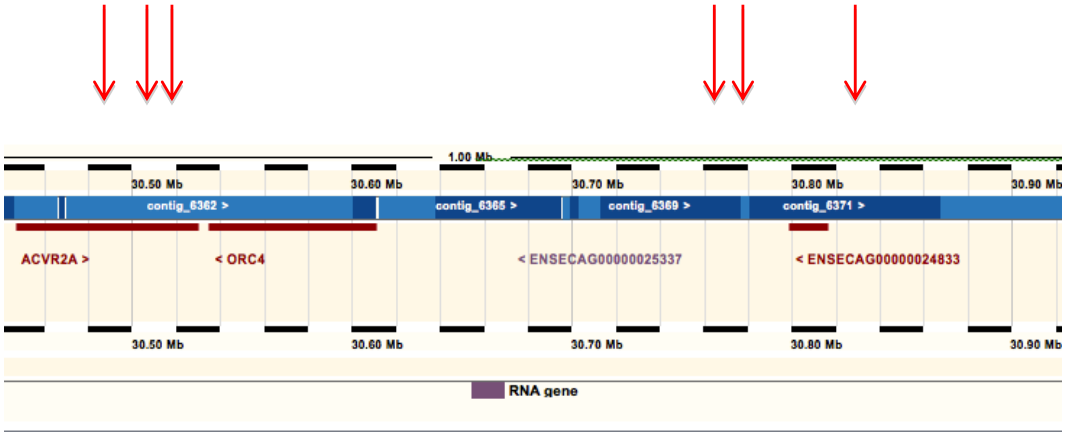


Figure 2. Ensembl. Equine chromosome 18: 30,442,277-31,038,477. Red arrows represent significant SNPs on GWAS analysis.

The six ECA18 SNPs span an interval of ~350 kb (349,806 base pairs). This interval of ECA18 contains two protein-coding genes, both expressed in equine skeletal muscle, activin A receptor, type IIA (*ACVR2A*) and origin recognition complex, subunit 4 like (*ORC4L*). PCR was attempted to amplify 5' and 3' untranslated regions (UTR), coding (exonic) regions and exon/intron boundaries (splice sites) from each gene to enable Sanger sequencing and detection of simple mutations (nonsynonymous substitutions, missense mutations and small insertions or deletions). Complete sequencing of these regions in *ORCL4* did not identify simple mutations associated with PSSM2. Sequencing of most *ACVR2a* exons and intron/exon boundaries has not identified simple mutations within these sequences. Due to poor annotation and assembly of the reference equine genome (Twilight)[55], amplification and sequencing of the 5" UTR/ promoter region and the first exon of *ACVR2a* was unsuccessful. Attempts to design primers and "walk" through the region with Sanger sequencing of genomic DNA in PSSM2 cases and controls and BAC clones (bacterial artificial chromosome clones)[56] have also been unsuccessful and suggest that large structural variants may be present in this region.

In addition, 3 other SNPs on chromosome 18, outside *ACVR2,* had significant p-values on the GWAS analysis. The region is approximately 68 Kb and is composed of several repetitive elements and contains no known genes. A novel protein coding 'mRNA' is present in this region (ENSECAG00000024833) and is also composed of several repetitive elements.

Limitations to the initial GWAS

The equine SNP50 chip has been successfully used in several association studies, but its performance can be limited by the lower SNP density and limited coverage in some regions of the genome. The markers originated from 7 equine breeds and have an inter-SNP space of 43 kb [57] . A newly develop equine SNP array has a total of 670,000 markers and approximately 8 SNPs per 50 kb window. The markers were sequenced from 166 horses representing 32 breeds and were further genotyped in 348 horses representing 10 major breed groups to assess SNP performance. The 670k markers were chosen based on a list of 2 million markers, from which haplotype information is available.

To increase the power and marker density of the initial GWAS performed to identify markers associated with PSSM2, cases and controls could be genotyped using the new 670,000 array. A second, more cost effective alternative to increase SNP density and therefore the power of the analysis is to predict genotypes by the use of imputation.

Research questions to be addressed for PSSM2

This section of the thesis study will address the genetic basis of type 2 PSSM by evaluating variants that could be associated with the disease.

Hypothesis and Specific Aims PSSM2

Identification of the genes and alleles underlying PSSM2

Rationale: Although the prevalence of the *GYS1* mutation is high in Draft Horses, approximately 25% of PSSM cases in Quarter Horses (QH) and 80% of PSSM cases in Warmbloods are not attributable to this mutation[9]. Histopathology and ultrastructural studies have demonstrated phenotypic differences between the *GYS1* (PSSM1) and non-*GYS1* (PSSM2) forms of PSSM[12]. Similar to PSSM1, there is evidence that PSSM2 is an inheritable disorder based on visual analysis of pedigree (Molly McCue, personal communication). The c*entral hypothesis* is that a gene or genes of moderate to major effect underlie type 2 PSSM in Quarter Horses. Preliminary work has identified a chromosomal region on equine chromosome 18 (ECA18) associated with PSSM2 in QHs, although the underlying mutation has not been identified. Thus the focus of Aim 2 is to thoroughly interrogate this region on ECA18, to identify any alleles segregating with PSSM2.

Aim 2: To thoroughly interrogate the genomic region on ECA18 associated with PSSM2, high throughput next-generation sequencing technology for genetic variants in PSSM2 cases and controls by target enrichment via long-PCR and next-generation sequencing of genomic DNA was used. This approach allowed me: 1) to overcome limitations in SNP-based fine structure mapping and Sanger sequencing of candidate genes imposed by haplotype structure and incomplete genome annotation in the horse; 2) to identify structural variants that may be missed with traditional Sanger sequencing technologies (i.e. large insertions, deletions, and/or inversions). Segregation of variants with disease status will be confirmed within the cohort used for our preliminary studies prior to validation in an independent population; 3) to interrogate the region on ECA18 where the novel protein coding 'mRNA' is present (ENSECAG00000024833), next generation whole genome sequence was performed on 3 type 2 PSSM horses and 3 controls using Illumina HiSeq, 100 bp paired end and 12x coverage.

The strength of association on ECA18 in PSSM2 Quarter Horses is moderate. Further, other work in our lab has demonstrated that Quarter Horses have very short haplotype lengths and has suggested that the 54k array maybe inadequate for association mapping in the Quarter Horse[57]. The inadequate marker density on the 54k array may

result in failure to identify causal loci, since SNP markers on the array may not be in strong LD with the causative alleles. To overcome the inadequate marker density in Aim 3, genotype imputation will be used to effectively increase marker density and the GWAS will be repeated to determine if there are additional loci associated with PSSM2.

Genotype Imputation and GWAS

The goal of genotype imputation is to predict the genotypes at the SNPs that are not directly genotyped in the study sample. A reference panel of haplotypes at a dense set of SNPs is used to impute into a study sample of individuals that have been genotyped at a subset of the SNPs. Genotype imputation can be carried out across the whole genome as part of a genome-wide association (GWA) study, increasing its power[58]. According to McCoy the size of the imputed population did not impact imputation success when assessing genotype imputation accuracy from the 54,000 (SNP50) to 65,000 (SNP60) equine genotyping array. Imputation success increased with larger reference population sizes and when imputed and reference populations were breed-matched. Linkage disequilibrium also influenced genotype accuracy. Breeds with longer LD (Thoroughbred) had higher imputation success than did those with shorter LD (QH)[59].

Since the 54,000 markers from the SNP50 are present in the new 670k equine SNP array and the haplotype information from the 1.8 million markers is available, imputing from 50k to 670K or 1.8 million markers is feasible. Imputing less dense genotype arrays is a cost-effective and powerful way to include genetic markers into association studies without the need to re-genotype the samples.

By identifying the genetic mutation that causes PSSM2, the sensitivity and specificity of diagnosing the disease improves, since there is controversy on PSSM diagnosis based on muscle biopsy evaluation. Also, we will be able to better understand the physiology of the disease and develop new therapeutic options and management recommendations.

Chapter 2

Normal equine skeletal muscle gene expression profiles prior to and after a standardized submaximal exercise trial

Summary

The objectives of this chapter are to identify the genes and transcripts expressed in normal equine skeletal muscle; and to determine the effects of a 3 week-period submaximal exercise training, and a bout of submaximal exercise on skeletal muscle gene expression. Middle gluteal muscle biopsies were collected from 6 normal horses prior to trial (t1) and before (t2) and after exercise (t3) following a 3-week controlled exercise protocol. RNA was extracted from muscle and RNASeq was performed (50 bp, single end reads).

90-95% of the reads mapped uniquely to the equine reference genome using Bowtie and Tophat. Approximately ~8,800 skeletal muscle transcripts with an FPKM greater than 2 were identified by Cufflinks. Read counts were generated by HTSeq and were used by EdgeR to measure gene expression profile differences between time points. This analysis revealed 528 genes differentially expressed with FDR lower than 0.05 between time point 1 (t1) and time point 2 (t2), and 108 genes differentially expressed between t2 and t3.

The top differentially expressed genes were imported into DAVID to extract information and group clusters according to their biological function. Several genes were identified as being important for the MAPK signaling pathway between t2 and t3 (effect of exercise on gene expression). Pathway analysis by GSEA (gene set enrichment analysis), which uses all genes in the dataset and their expression level revealed enrichment in pathways involved in glycolysis, mitochondrial biogenesis, muscle contraction and development and aerobic respiration.

Better annotation of the equine reference genome and manual annotation of the pathways identified by GSEA will provide more insights into the changes in gene expression that occur in normal equine muscle in response to submaximal exercise training. Also, de novo assembly of the reads will make the discovery of new transcripts

and isoforms possible and will provide a more complete picture of the equine muscle transcriptome.

Introduction

Descriptions of the skeletal muscle transcriptome and the effect of exercise on muscle gene expression have been extensively reported in humans[60] but only a few studies have evaluated equine muscle transcriptome and gene expression profile changes in response to an exercise trial. All muscle gene expression studies in horses so far have been performed in the Thoroughbred (TB) [61 62]. In one study, gene expression measured in the triceps brachii muscle by RNASeq before and after 30 minutes of trotting in six retired TB horses showed changes in genes previously reported to affect racing performance (*H1F1; ADRB2; PPARD; VEGF; TNC; BDNF*) and exercise training (*ACTR3B; FBXO32; PER3; C1orf51; GATM)* [61]. By using digital gene expression (DGE) in the gluteal muscle of seven, 2 year-old TB horses before and after 10 months of training, overrepresentation of mitochondrial genes and genes involved in muscle contraction and aerobic respiration, as well as high expression of creatine kinase (*CKM*), mitochondrial creatine kinase (*CKMT2)* and *ACTA1* were observed[62]. Only 25% (5,068 unique genes) of annotated equine genes could be evaluated using the DGE technique. Overall, metabolic changes in response to exercise are related to increased oxidative enzyme activity and oxidative fibers, and an increase in volume density of mitochondria. Horses performing at submaximal aerobic exercise intensities benefit from enhanced delivery of oxidative substrates to muscle fibers as well as improved oxidative metabolism of glycogen and FFA[40].

None of the studies described above have evaluated the effect of controlled regular submaximal exercise in horses in gluteal muscle biopsies using RNASeq. By using RNASeq, genes expressed in equine muscle and their changes in expression values with exercise can be globally evaluated. RNA-seq has a higher sensitivity and lower technical variation, in comparison to microarray. The Illumina sequencing data are highly replicable and the information in a single lane appears comparable to that in a single array in enabling identification of differentially expressed genes, while allowing for detection of low-expressed genes, alternative splice variants, and novel transcripts [63].

The objectives of this chapter are to identify the genes and transcripts expressed in normal equine skeletal muscle; and to determine the effects of a 3 week-period submaximal exercise training, and a bout of submaximal exercise on skeletal muscle gene expression.

Material and Methods

Study Design

Horses. Sample sizes needed to detect gene expression differences were estimated by n > (σ2 log f)/ (z2 1-α), (σ2, variance in gene expression was estimated from data in other species[64], fold change = 2; and α=0.05). The sample size was determined to identify the changes between groups in chapter 3. Based on this calculation, a minimum of 5 biologic replicates are necessary to detect differences in gene expression, consistent with standard recommendations[65]. Six healthy control QHs; 4 mares and 2 geldings ranging in age from 1.5 to 14 years were utilized for the study. Horses had normal muscle histology, normal amount of glycogen in the muscle, normal serum CK activities and no history of rhabdomyolysis. Prior to training, horses were unfit, and had a minimum of 90 days rest in a small paddock with no forced exercise.

Collection of biopsy samples and training protocol. Baseline percutaneous gluteal muscle biopsy samples were collected for all horses prior the start of treadmill exercise[20] (time point **t1**). Biopsy samples consisted of at least 0.25 g of tissue from either the left and right gluteal muscle; samples were immediately frozen in liquid nitrogen and stored at -80˚C pending further analysis. Horses had up to a two-week period of acclimatization to treadmill training (**Figure 3**). After the acclimatization horses completed a 3-week training period. Daily exercise consisted of 4 minutes of walk, and then alternating intervals of 2 min walk (1.9 m/s) and 2 min trot (3.0 – 3.8 m/s) for 20 min. At the end of the training period (day 21), muscle biopsy samples were collected before (time point **t2**) and after submaximal exercise (time point **t3**).

RNA extraction and sequencing

Total RNA was isolated from muscle tissue using Trizol reagent, quantified by Nanodrop spectrophotometry, and examined for integrity of RNA bands on an Agilent Bionalyzer. Library construction was performed using the Illumina HiSeq genome analyzer at the University of Minnesota Genomics Center (UMGC), using the TruSeq RNA V2 LT Illumina kit. First, the procedure purifies messenger RNA from total RNA, using polyA selection. The mRNA was then chemically fragmented and converted into single-stranded cDNA using random hexamer priming. Next, the second strand was generated to create double-stranded cDNA for the TruSeq library construction workflow, where adapters are ligated to ends of the fragments. Each library was sequenced (50 bp single end reads) at a rate of 40 million reads per sample. Individual libraries were sequenced in more than one lane so possible batch effects could be identified later[66].

Quality Control and Mapping

Fastq files originating from sequencing were converted to fastq Sanger format. Prior to mapping, quality control of the reads was performed using FastQC program. QC reports were generated for every library and analyzed for per base sequence and per sequence quality score, per base and per sequence GC content, per base N content, sequence length distribution, duplication levels and overrepresented sequences and Kmer content. Reads with a quality score lower than 20 were trimmed. The reads were then mapped to the equine reference genome (EqCab2) using Bowtie2 and the splice-aware aligner Tophat2[67]. First, Tophat uses Bowtie to align reads to the reference and breaks sequences that Bowtie cannot align on its own into shorter segments. Bowtie does not allow large gaps between read alignments to the reference genome, so it does not align reads that contain introns. The short segments created by Tophat[67,68] are then aligned to the reference genome. When segments created from the same reads aligned to the genome far from each other, Tophat infers that the reads span a splice junction and builds an index of splice sites in the transcriptome. A gene model annotation file (GTF; gene transfer format) from ensemble was provided for the analysis. A GTF file is a tab-delimited file describing genomic features and their location. The reads from the same individual (library) were mapped by lane and the generated BAM files were then merged

for follow-up analysis. A BAM file is a binary version of a SAM file, which is a tab-delimited text file that contains sequence alignment data.

Transcript assembly and normalized gene count

Transcript assembly and abundance estimation was then performed by Cufflinks [69](version 0.0.6) with the previously supplied GTF file. Cufflinks is a reference-based transcriptome assembler. It assembles individual transcripts from RNA-Seq reads that were aligned to the reference genome. Each sample (library) was assembled individually. After assembly, Cufflinks calculates gene expression abundance by normalizing the counts into RPKM levels (reads per kilobase per million). RPKM=10^9*C/NL; where C=number of mapped reads that fall into the gene's exons; N=total number of mapped reads in the experiment; and L=sum of the exons in base pairs. The program outputs a list of all the genes expressed in muscle and their respective RPKM values[70].

Non-normalized (raw) gene count

A non-normalized gene count (raw count) was also performed, as this is the required input for differentially expression analyses based on negative binomial distribution (edgeR). The BAM files originated from mapping with Tophat were sorted and indexed by Samtools. HTSeq-count (a python package) was used to count the reads that mapped to each one of the equine genes present in the GTF file. Given a file with aligned sequencing reads and a list of genomic features, HTSeq counts how many reads map to each feature. In this case, the read counts are the total number of reads aligned to each gene. HTSeq has three overlap resolution modes that dictate how aligned reads that overlap more than one genomic feature are treated (union, intersection-strict, intersection-nonempty - **figure 4**). A union counting mode was used to count the reads. The output is a tab-delimited table of read counts for each gene listed in the genomic annotation set.

Differentially expression analysis - Edge R

Gene differential expression analysis was performed on edgeR (Bioconductor R package[71]) based on negative binomial distribution since there is more variation in RNA-seq data that can be accounted for by the Poisson model. In RNA-seq data, the variance

grows faster than the mean, which is known as overdispersion, so the mean and the variance are modeled separately. EdgeR works on a table of read counts where rows correspond to genes and columns to independent libraries containing the non-normalized count produced by HTSeq. This design matrix was created into an object (DGE list) in R. Adjustment for sequencing depth is always done automatically by EdgeR, since different libraries can have different sequencing depths and it can affect read counts.

Another important issue is dealing with the total RNA output in a flow cell. Genes that have a very high expression level can consume most of the library size and cause the remaining genes to be under sampled, which can make them look down regulated unless this effect is controlled for in the analysis. The calcNormFactor function normalizes for RNA composition using a trimmed mean of M-values (TMM) between each pair of samples. This creates the effective library size, which is the dataset used for downstream analysis. EdgeR considers that GC content and gene length have little effect on differential expression analysis since they do not change from sample to sample (the same genes will have the same length and GC content regardless of the sample). Then dispersion was estimated and a linear model fitted (GLM) to ask the questions of interest (**figure 5**).

Pathway Analysis

David

This approach focuses on examining a preset list of genes (which have the largest difference between groups), to evaluate their biological function. Following identification of genes differentially expressed, biologic themes, disease-gene relationships and metabolic pathways within the gene lists was identified with tools available from the DAVID bioinformatics resource (NIAID, http://david.abcc.ncifcrf.gov/). DAVID consists of an integrated biological knowledgebase and analytic tools aiming at systematically extracting biological meaning from large gene lists. This is more of an exploratory procedure than a pure statistical solution. The goal of the analysis is to map a large number of interesting genes in a list to the associated biological annotation, and then

statistically highlight the most over-represented (enriched) biological annotation out of thousands of linked terms and contents.

First a gene list is provided for the analysis (differently expressed genes). In the gene name batch viewer, a gene history report is provided and information about each gene can be accessed. The genes from the gene list are then clustered into groups according to their biological function. An enrichment score is provided and represents how important this gene group is in the gene list (the higher the level the greater the importance). Next, the functional annotation chart tool identifies the most relevant (over-expressed) biological terms with a given gene list (40 annotation categories are used). In this output, enriched functional annotations are divided into different categories which includes term (e.g.. transcription factor activity), number of genes in each category, percentage of genes involved is this category from the total gene list and a p-value (modified Fisher exact p-value. Fisher Exact test is adopted to measure the gene-enrichment in annotation terms. It ranges from 0 to 1. A p-value of zero represents perfect enrichment). Other tools are also available to visualize the same data in different formats (pathway map viewer, tabular format and others). Clusters with FDR < 0.05 were considered significant.

Gene Set Enrichment Analysis – GSEA

Traditional strategies for gene expression analysis have focused on identifying individual genes (single gene analysis) that exhibit differences between two states of interest. Although useful, they fail to detect biological processes, such as metabolic pathways and responses, that are distributed across an entire network of genes and subtle at the level of individual genes[72]. Cellular processes often affect sets of genes acting together. An increase of 20% in all genes encoding members of a metabolic pathway may dramatically alter the flux through the pathway and may be more important than a 20-fold increase in a single gene[72].

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (cases and controls for example). Gene sets are defined based on prior biological knowledge about the physiology of the disease of interest. The priori defined set of genes is chosen by the user, based on what is known about the pathological process of interest. It is a tool to interpret gene expression profile results, extracting meaning from their expression level and gaining insights into biological mechanisms[72]. Instead of a single gene analysis, GSEA calculates expression data at the level of gene sets. The goal of the program is to determine whether members of a gene set tend to occur at the top or bottom of the expression dataset, in which case the gene set is correlated with the phenotype of interest. Given an *a priori* defined set of genes *S* (e.g., genes encoding products in a metabolic pathway, located in the same cytogenetic band, or sharing the same GO category), the goal of GSEA is to determine whether the members of *S* are randomly distributed throughout *L* or primarily found at the top or bottom. We expect that sets related to the phenotypic distinction will tend to show the latter distribution. GSEA applies Kolmogorov-Smirnof test to find asymmetrical distributions for defined blocks of genes in datasets whole distribution.

The program requires an expression dataset (created by the user). The first column contains information about genes (HUGO gene names in case of RNA-Seq data), the second column a description of the genes that is ignored by the program, followed by sample names (third column onwards). The rows are filled with expression value for each feature (gene) in each sample. Each column contains expression values for 1 sample. A phenotype label set (associates each sample to a phenotype) and gene sets that are exported from MSigDb or created by the user are also required. The expression data used in the present study was the read count number generated from HTSeq.

First GSEA calculates a rank for each one of the genes present in expression dataset that are also present in the gene set list from a specific predefined pathway (list of genes expressed in a specific condition and their expression values). The genes are ranked based on their correlation with the phenotype and a ranked list is created (list L). Then an enrichment score (ES) is calculated and reflects the degree to which a set S (gene list from a specific pathway) is overrepresented at the extremes (top or bottom) of the

entire ranked list L. The score is calculated by walking down the list L, increasing a running-sum statistic when we encounter a gene in *S* and decreasing it when we encounter genes not in *S.* The magnitude of the increment depends on the correlation of the gene with the phenotype. The enrichment score is the maximum deviation from zero encountered in the random walk; it corresponds to a weighted Kolmogorov–Smirnov-like statistic. Next GSEA estimates the significance level of the ES by using an empirical phenotype-based permutation test procedure that preserves the complex correlation structure of the gene expression data. Estimated significance level to account for multiple hypothesis testing is also performed since an entire database of gene sets is evaluated. A normalized enrichment score (*NES*) is also performed to account for the size of the set, which allows comparisons between pathways. The proportion of false positives is then calculated using false discovery rate (FDR) corresponding to each *NES*. The FDR is the estimated probability that a set with a given *NES* represents a false positive finding; it is computed by comparing the tails of the observed and null distributions for the *NES*. The top portion of the plot shows the running ES for the gene set as the analysis walks down the ranked list. The score at the peak of the plot (the score furthest from 0.0) is the ES for the gene set. Gene sets with a distinct peak at the beginning or end of the ranked list are generally the most interesting.

The leading edge subset of a gene set is the subset of members that contribute most to the ES. Leading-edge subset are the genes in the gene set *S* that appear in the ranked list *L* at, or before, the point where the running sum reaches its maximum deviation from zero. The leading-edge subset can be interpreted as the core of a gene set that accounts for the enrichment signal.

The Molecular Signature Database (MSigDB) is a catalog of 1,325 gene sets, which are annotated gene sets for use with GSEA software. From this web site, the user can search for genes by keywords, browse gene sets by name or collection, examine a gene set and its annotation and download a gene set for analysis. For this study, gene sets for pathway analysis from the MSigDB database were selected based on general key words known to be important for training and exercise effect and include aerobic respiration and mitochondrial function, muscle contraction and structure, fat oxidation, glycogen and glycolysis [62].

A workflow representing the RNASeq analysis is represented in **Figure 6.**

Results

RNA and sequencing quality controls

Total RNA integrity following isolation using an Agilent Technologies 2100 Bioanalyzer revealed an RNA Integrity Number (RIN) value greater than eight for all samples. An average base calling Q score (a PHRED score) was above the quality cutoff of 30 (ie < 1/1000 probability that the base call is incorrect) in all lanes. All samples met the sequencing criteria of 40 million reads/sample.

Mapping summary statistics

Approximately 90-95 % of the total reads per library mapped to the equine reference genome and approximately 8,800 transcripts were identified in equine muscle with FPKM greater ≥ 2. Approximately 8% of the reads on each library did not align uniquely to the reference.

List of genes expressed in normal equine muscle and FPKM values - Cufflinks

Time point 1 (t1) - pretrial

For time point 1 8949 transcripts with FPKM values greater than 2 and 9782 transcripts with FPKM values greater than 1.5 were identified using Cufflinks. The top 100 assembled transcripts and their FPKM values for each horse are shown on **table1**.

Time point 2 (t2) – pre-exercise

For time point 2 9074 transcripts with FPKM values greater than 2 and 9966 transcripts with FPKM values greater than 1.5. The top 100 assembled transcripts and their FPKM values for each horse are shown on **table 2**.

Time point 3 (t3) – post-exercise

For time point 3 8839 transcripts with FPKM values greater than 2 and 9676 transcripts with FPKM values greater than 1.5 were identified using Cufflinks. The top 100 assembled transcripts and their FPKM values for each horse are shown on **table 3**.

Differential expression analysis – Effect of exercise and training

Differential expression analysis performed between healthy horses in all 3 time points revealed 528 genes differently expressed with FDR lower than 0.05 between time point 1 (t1) and time point 2 (t2) (**table 4**) and 108 between t2 and t3 (**table 5**). ID conversion from ensemble gene names to HUGO gene names was performed on DAVID and BIOMART. Hundreds of genes were not annotated in the equine reference genome and were therefore manually annotated using blast. Not all genes could be manually annotated. From those comparisons, the ones of major interest are between t1 and t2 (effect of training on gene expression) and t2 and t3 (effect of exercise on gene expression).

Biologic clustering of differentially expressed genes

The top differentially expressed genes between time points 1 and 2 (training effect), and time points 2 and 3 (exercise effect), were clustered by DAVID based on their biological function. Inflammation and immune reaction biologic pathways were overrepresented in genes that were differentially expressed between time point 1 and 2 (effect of training). After exercise (between time point 2 and 3), overrepresented biologic processes included several pathways involved in up-regulated gene transcription in general as well as MAPK pathways (**table 6**).

Gene Set Enrichment Analysis

Table 7 summarizes all the key words and pathways from the MSigDB database used for GSEA analysis. The two curated pathways involved in glycogen metabolism and breakdown identified in the database (keyword glycogen) were tested for enrichment

between t1 and t2 and t2 and t3. No significant enrichment was observed between those time points. Two pathways were identified using the keyword glycolysis. When testing for enrichment score, time point 2 showed enrichment of glycolytic pathways in comparison to t1 and t3. Fatty acid oxidation search revealed 2 pathways of interest, but no significant enrichment was observed between normal horses at different time points. Using the keyword aerobic respiration identified one pathway. No difference apart from suggestive of enrichment on t3 in comparison to t2 was observed (pathway: cellular respiration). Evaluation of pathways involved in mitochondrial biogenesis revealed enrichment in t2 in comparison with t1. Enrichment in the muscle contraction and development pathways was also observed in t2 in comparison to t1 (after training) and t3 in comparison with t2 (after exercise) (**table 8**).

Discussion

More than 8000 transcripts were identified in equine muscle. Genes involved in respiratory oxidation and glycolysis were enriched after sub-maximal exercise training, as were pathways involved in mitochondrial biogenesis and muscle contraction.

High RNA integrity, high mapping quality scores and a high sequence depth made it possible to map reads to the genome at a high mapping percentage. The reads created by the library process were short which can make it difficult to map it with certainty to the genome, but only 8% of the reads per library did not map uniquely. Although most of the reads mapped uniquely, TopHat is able to more accurately discover splice junctions with longer (75 bp or longer) and paired end reads.

Approximately 8,000 genes were expressed in equine muscle with an FPKM > 2. Analysis of the genes differentially expressed between time points using DAVID identified dozens of biologic annotations that were enriched within the differentially expressed (DE) gene set. The vast majority of the biologic functions that were overrepresented in the DE gene set after 3 weeks of training were involved in

inflammation and immune reaction pathways. The biologic processes that were identified as enriched after exercise included several pathways that indicated up-regulated gene transcription in general as well as a few pathways that can be tied to response to exercise such as the MAPK pathway. While using edge R evaluates the statistical significance of genes by itself, there are advantages to analysis set of genes together according with their biological function. GSEA considers all of the genes in an experiment, not only those above an arbitrary cutoff in terms of significance and evaluates their significance as a set. Evaluation of GSEA of the genes differently expressed between t1 and t2 and t2 and t3 revealed enrichment on respiratory oxidation, glycolysis, mitochondrial biogenesis and muscle contracture and development. No enrichment in fatty acid oxidation was observed between time points.

The study where gene expression profiles were measured in blood and the triceps brachii muscle in 6 retired thoroughbred horses before and after exercise found 1,285 genes differentially expressed genes (62 up and 80 down regulated in blood and 878 up and 285 down regulated in muscle). The samples were collected before and right after 30 minutes of trotting and expression was measured by RNASeq. According to the author, six genes previously associated with racing performance (*H1F1; ADRB2; PPARD; VEGF; TNC; BDNF*) and five genes previously associated with exercise training (*ACTR3B; FBXO32; PER3; C1orf51; GATM*)[61] were found to be upregulated after exercise. In the dataset presented in this chapter only genes *FBXO32 (*F-Box Protein 32)*, C1orf51* (CIART; circadian associated repressor of transcription) and *BDNF (*brain-derived neurotrophic factor*)* were found to have a significant difference in expression (FDR <0.05) between time points 1 and 2 (effect of training). Those 3 genes had a higher expression at t2 in comparison to t1 (up regulation in response to training). The fact that the other genes were not differentially expressed on our dataset could possibly be explained by the different type of exercise that the horses performed in both studies, the different muscle group that was sampled and the different age and breed of the horses. It is also possible that the genes reported in the present study are not truly associated with performance in thoroughbred horses.

In one of the published studies, digital gene expression (DGE, Illumina) profiling was used in seven two-year-old untrained Thoroughbred race horses to characterize genes expressed in equine skeletal muscle and to identify the subset of genes that were differentially expressed following a ten month period of exercise training. The horses were raised on the same farm for the previous 2 - 3 months, and destined for flat racing. All horses undertook a regular exercise regime with the same trainer for 10 months (light canter once a day six times a week and higher intensity exercise no more than once a week which consisted of warm-up (walk and trot) followed by gallop with velocities reaching maximal intensity for 800-1,000 m. Gluteus medius muscle biopsies at 2 time points (at rest: untrained and trained) were performed. Using DAVID an overrepresentation of mitochondrial genes, and genes involved in muscle contraction and aerobic respiration was observed[62]. Creatine kinase (*CKM*) was the most abundant expressed gene in equine muscle, as well as *CKMT2* and *ACTA1* (skeletal muscle actin, alpha 1). 92 transcripts were differentially expressed (FDR=0.05) after 10 months of training. Genes with higher post-training basal mRNA levels included those involved in the mitochondria, ubiquitination and circadian rhythm regulation, whereas genes with significantly reduced mRNA were mainly associated with cytoskeletal structure and the control of growth and development. Also, following exercise training an increase in genes involved in oxidative phosphorylation was observed. Functional analysis of all expressed genes using FatiScan revealed an asymmetric distribution of 482 Gene Ontology (GO) groups and 18 KEGG pathways. Functional groups displaying highly significant ($P < 0.0001$) increased expression included mitochondrion, oxidative phosphorylation and fatty acid metabolism while functional groups with decreased expression were mainly associated with structural genes and included the sarcoplasm, laminin complex and cytoskeleton. Several findings in this study were similar to what we describe in this chapter, including changes in respiratory oxidation, mitochondrion and muscle structure. The differences between studies again are likely due to the intensity of the exercise performed and the duration of the trial. Also, one of the limitations of DGE technique is the number of genes that could be evaluated for expression since it relies on an NlaIII restriction site. Seventy eight percent (n= 19,271) of currently annotated equine genes are quantifiable using DGE, but only the 20% of reads that unambiguously

matched Ensembl genes were used for further analysis, which represented 5,068 unique genes (~25% of annotated equine genes).

One of the limitations of the study presented in this chapter is the fact that hundreds of transcripts that were found as being expressed did not have gene IDs and had only ensemble IDs. Several of those IDs were converted using Biomart and DAVID, but a large number had to be manually annotated. Also, to perform GSEA the equine transcript names have to be converted to HUGO gene names, which can be a challenge since a greater number of isoforms are present in the human database and the conversion is not always straightforward. A reference genome with better annotation and the use of a more consistent gene nomenclature between species would greatly improve this issue. Further, TopHat and Cufflinks are reference-based aligner and assembly tools, so the transcripts and isoforms identified are highly dependent on proper assembly and annotation of the reference genome; therefore it is quite possible that genes or novel transcripts/isoforms were missed by these tools. *De novo* assembly of the reads would also enable a more complete description of the equine muscle transcriptome. In model species, like mice and in humans with well-annotated genomes, missing data due to poor annotation is less of a problem; however it still poses a significant challenge in species where the reference is not as well annotated like the horse. *De novo* assembly of the raw reads does not rely on a reference genome, allowing the issue of poor annotation to be partially overcome and allowing for identification of novel transcripts and isoforms.

It has been shown that gene expression in muscle can increase hours after exercise. Several myogenic and metabolic genes, for example, have a peak in gene expression 4 to 8 hours post-exercise[73]. Our samples were collected immediately after exercise, so it is possible that some genes would have a higher level of expression if the biopsies were performed in different time points. Although this might be a limitation of the study, muscle biopsy immediately after exercise enables the identification of differentially expressed early response genes that are rapidly induced by exercise.

In conclusion ~8000 transcripts were found to be expressed in equine muscle. Once de novo assembly of the reads is performed, the number of transcripts and isoforms will likely increase significantly. When comparison of the gene expression profile was made between time points, genes involved in respiratory oxidation and glycolysis were found to be enriched after sub-maximal exercise training, as well as pathways involved in mitochondrial biogenesis and muscle contraction. Further annotation of available pathways and better annotation of the reference genome would probably allow the discovery of other important pathways.

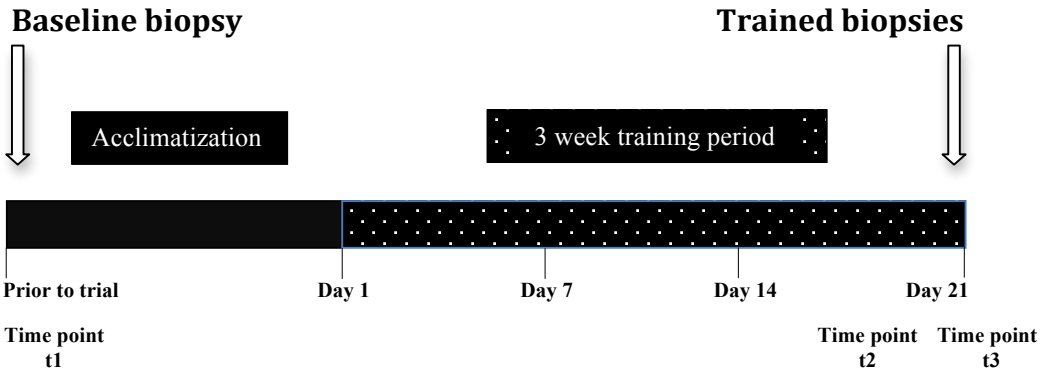


Figure 3. Sample collection protocol. Horses were acclimatized to the treadmill for 1 to 2 weeks. Horses trained 5 days/week for 3 weeks during the training period. On day 21 biopsies were taken before and after exercise

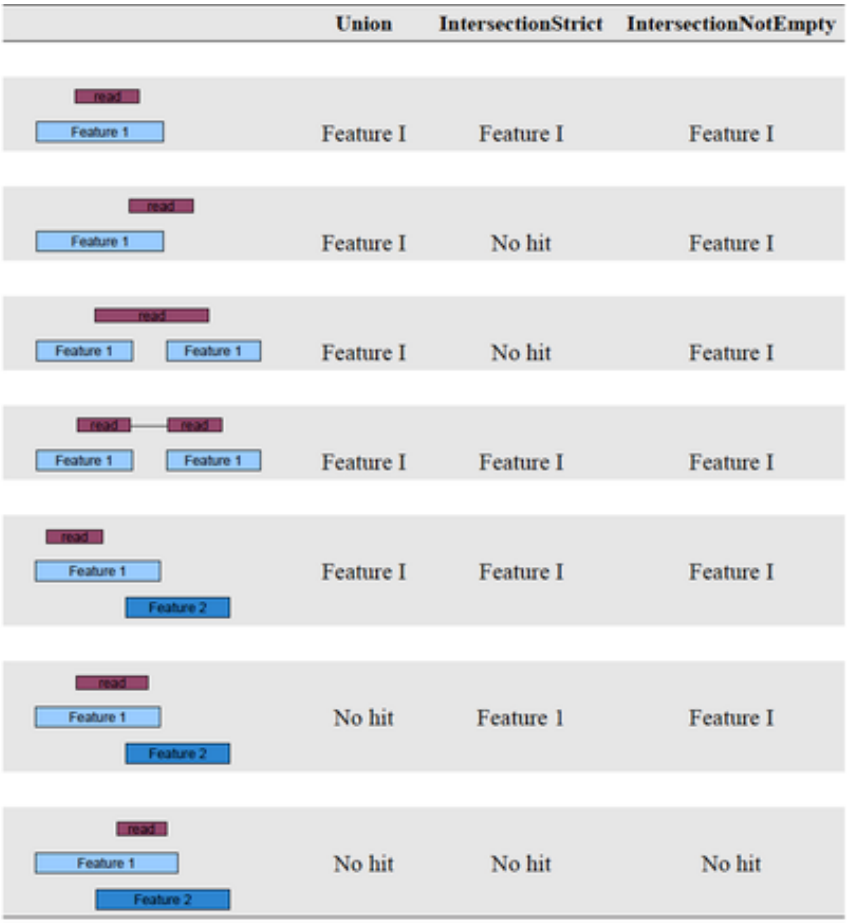


Figure 4. The three overlap modes of HTSEq.
*image from genomic-ranges bioconductor package:
http://www.bioconductor.org/packages/release/bioc/html/GenomicRanges.html

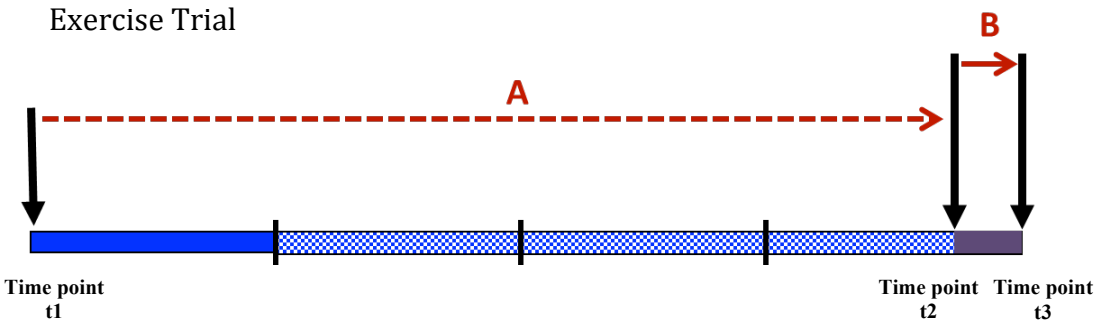


Figure 5. Comparisons performed between time points. A. Training effect within group. B. Exercise effect within group.

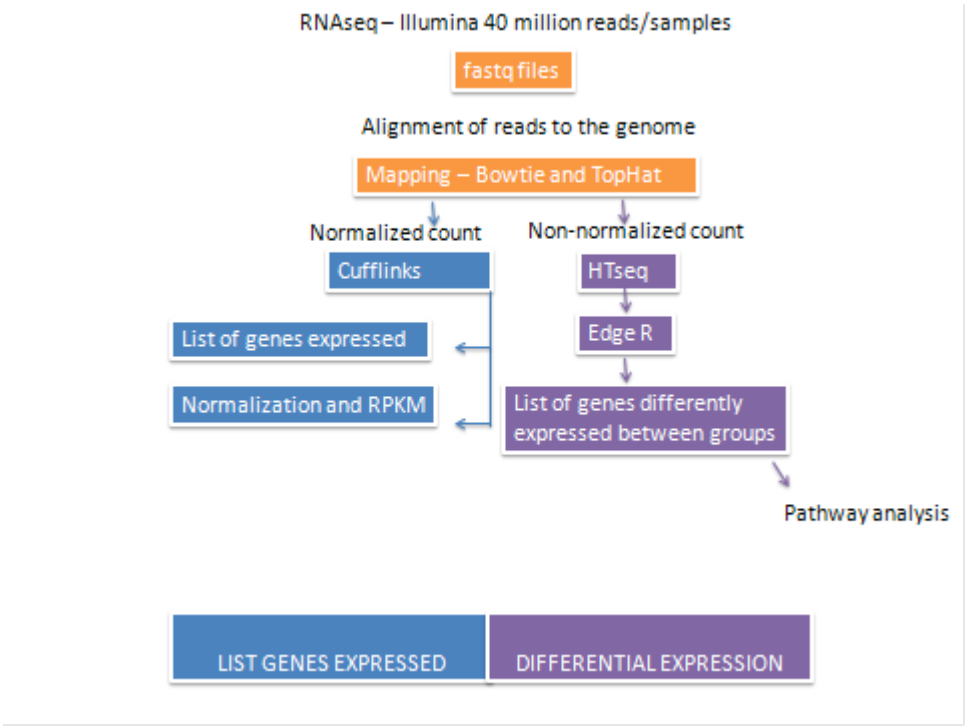


Figure 6. Workflow of RNASeq data analysis.

Table 1. Output from Cufflinks. Genes expressed in normal equine muscle at time point1. **The top 100 genes with the highest FPKM are shown on the table**

Ensembl Transcript ID	Ensembl Gene ID	Gene	FPKM 1	FPKM 2	FPKM 3	FPKM 4	FPKM 5	FPKM 6	Average
ENSECAT00000029843	ENSECAG00000027698	Novel Mt tRNA	236985	389984	396414	458878	154167	213565	308332.1667
ENSECAT00000005510	ENSECAG00000005606	SLN	152687	136560	187663	158261	184796	158028	162999.1667
ENSECAT00000029864	ENSECAG00000027702		43612.5	81824.3	71112.4	81895.3	44225.2	57060.2	63288.31667
ENSECAT00000029837	ENSECAG00000027692	MT-CO2	35241.1	47652.7	38674.8	40270.5	29445.9	38704.7	38331.61667
ENSECAT00000021003	ENSECAG00000019728	TNNC2	36028.6	32120	40642.5	38206.4	38579.3	38574.6	37358.56667
ENSECAT00000027340	ENSECAG00000025328		65971.8	28645.4	26957.3	20200.3	41367.2	33012.8	36025.8
ENSECAT00000029850	ENSECAG00000027674		28218.4	44570.4	33396.4	37085.9	15294	20134	29783.18333
ENSECAT00000029222	ENSECAG00000027056		50681.5	22346.8	18171.3	13389	30865.4	22114.2	26261.36667
ENSECAT00000029835	ENSECAG00000027695	ND3	21292.7	29873.3	21886.9	19042.2	23030.8	28224.3	23891.7
ENSECAT00000029790	ENSECAG00000027624		45819.9	17638.6	16592.7	11694.1	24379.7	20419.9	22757.48333
ENSECAT00000018609	ENSECAG00000017437	MYLPF	22008.1	18659.7	27409.5	23190.9	21287.7	23351.8	22651.28333
ENSECAT00000016196	ENSECAG00000015494	TNNI2	19698	15720	20625.6	17139.3	18968.9	17977.6	18354.9
ENSECAT00000029690	ENSECAG00000027524		32276.2	12912.9	13029.9	10991.7	19016.7	14265.1	17082.08333
ENSECAT00000000176	ENSECAG00000000207	ACTA1	19100.6	17923.6	20563.3	10675.3	14933.4	17738.7	16822.48333
ENSECAT00000029856	ENSECAG00000027684	MT-ND1	13858.9	23330	13794.5	14944.2	16453.5	17989.3	16728.4
ENSECAT00000029845	ENSECAG00000027669	MT-CYB	13552.4	19576.2	14954.2	18096.7	15841.7	16799	16470.03333
ENSECAT00000029858	ENSECAG00000027675	MT-ND4	13656.5	20283.9	15220.9	18935.2	13336	14558	15998.41667
ENSECAT00000029841	ENSECAG00000027677	ND6	14383.8	0	22344.5	24612.3	14773	17916.7	15671.71667
ENSECAT00000029834	ENSECAG00000027681	ND2	12514.8	23046.1	11943.6	14064.4	13997.3	17868.6	15572.46667
ENSECAT00000023921	ENSECAG00000022433	CKM	15460.7	18654.9	19482	15284.8	22748.7	0	15271.85
ENSECAT00000006002	ENSECAG00000005145	MYL1	14153.2	10335.4	20236.9	11255.2	14871.7	15834.6	14447.83333
ENSECAT00000029852	ENSECAG00000027682		13336.4	20838.3	13002.3	3422.51	14419.7	20444.7	14243.985
ENSECAT00000025982	ENSECAG00000023971	TNNT3	14672.3	12203.3	16775.7	13372.8	12507	13807.6	13889.78333
ENSECAT00000004791	ENSECAG00000003121	ALDOA	0	16700.4	18686.8	17484.4	23513.3	0	12730.81667
ENSECAT00000017356	ENSECAG00000016536		33757.9	580.021	38334.3	925.578	493.022	469.943	12426.794
ENSECAT00000029847	ENSECAG00000027680		10562.6	13990.2	14424.8	15006.6	6749.67	13373.5	12351.22833
ENSECAT00000029836	ENSECAG00000027670		9325.68	12684.2	14003.6	13089.9	6451.36	12074.9	11271.60667
ENSECAT00000022265	ENSECAG00000020981		9500.11	8173.69	8530.74	9254.03	8943.97	8996.02	8899.76
ENSECAT00000023721	ENSECAG00000022051	G3PDH	10113.4	7092.29	8730.72	7174.73	9221.53	10202.5	8755.861667
ENSECAT00000029862	ENSECAG00000027694	MT-ND5	8411.54	0	10893.2	12288.2	9321.68	11098.4	8668.836667
ENSECAT00000018976	ENSECAG00000017982	MB	9477.82	9502.3	8891.69	6621.11	7996.64	7851	8390.093333
ENSECAT00000019534	ENSECAG00000018348	TPT1	7950.77	6504.03	8073.62	6994.7	7511.26	8940.54	7662.486667
ENSECAT00000015761	ENSECAG00000014949	TPM1	5770.12	4944.28	7287.16	7059.05	6043.99	7374.24	6413.14
ENSECAT00000029848	ENSECAG00000027667		5308.86	6198.24	7158.24	8457.57	5177.37	6164.33	6410.768333
ENSECAT00000029839	ENSECAG00000027668		4840.31	6230.39	6624.17	6254.98	2797.55	5478.13	5370.921667
ENSECAT00000000003	ENSECAG00000000008	RPLP2	4518.02	4626.21	5297.79	5865.1	4631.94	5575.37	5085.738333
ENSECAT00000029015	ENSECAG00000026987	NDUFA4	5217.28	4949.5	4493.55	4713.51	4831.66	4692.72	4816.37

ENSECAT00000029844	ENSECAG00000027687		3635.76	5043.67	4661.4	5202.53	3313.01	5688.85	4590.87
ENSECAT00000021682	ENSECAG00000020400	ENO3	5123.24	3632.03	4850.18	3930.21	5088.74	4788.87	4568.878333
ENSECAT00000008746	ENSECAG00000008542		4991.39	3516.15	4262.49	3472.44	5566.78	5288.59	4516.306667
ENSECAT00000018208	ENSECAG00000017292	ATP5I	4426.93	3724.07	3778.84	4048.75	3443.96	3605.79	3838.056667
ENSECAT00000029854	ENSECAG00000027690		3118.09	3386.23	5163.36	3686.67	1542.64	3608.01	3417.5
ENSECAT00000022404	ENSECAG00000020669	LDHA	3882.95	2891.62	4069.56	2775.7	3247.17	3507.72	3395.786667
ENSECAT00000007368	ENSECAG00000007378	C14orf2	3552.85	3337.8	3142.78	2987.58	2955.83	3387.32	3227.36
ENSECAT00000007630	ENSECAG00000007612	PPP1R1A	3355.14	2492.49	3123.89	3495.03	3502.35	3131.45	3183.391667
ENSECAT00000005912	ENSECAG00000005539	PGAM2	3226.54	2463.18	3609.59	2879.31	3223.38	3446.05	3141.341667
ENSECAT00000029543	ENSECAG00000027377	5_8S_rRNA	7778.62	1468.5	2834.59	1448.31	780.624	4222.12	3088.794
ENSECAT00000014627	ENSECAG00000013777	YBX3	2950.36	2658.99	2733.07	3441.27	3575.22	3029.79	3064.783333
ENSECAT00000009354	ENSECAG00000009190	TMSB4	3097.35	3293.83	3145.07	2615.06	2352.6	3382.75	2981.11
ENSECAT00000026040	ENSECAG00000024000	PYGM	2596.77	2795.47	2855.93	3360.38	3107.5	2700.24	2902.715
ENSECAT00000029842	ENSECAG00000027673		2251.55	3807.48	3164.9	3706.63	1725.37	2109.57	2794.25
ENSECAT00000029838	ENSECAG00000027678		1678.91	3630.75	2716.75	3305.35	1547.15	2872.04	2625.158333
ENSECAT00000017584	ENSECAG00000016730		1101.81	4829.49	1036.45	1387.83	3129.95	3359.25	2474.13
ENSECAT00000016514	ENSECAG00000015581	TPI1	2500.07	1914.21	2797.6	2390.13	2496.21	2479.27	2429.581667
ENSECAT00000017694	ENSECAG00000016739	MYOZ1	2442.56	2141.76	2304.88	2404.77	2262.22	2292.65	2308.14
ENSECAT00000028889	ENSECAG00000026905	RPL37	2048.21	1648.43	2155.98	1743.77	2403.85	2353.26	2058.916667
ENSECAT00000022182	ENSECAG00000020407	MYBPC2	2150.03	1773.7	1592.04	2246.28	2180.92	2021.29	1994.043333
ENSECAT00000018307	ENSECAG00000017154		1787.3	1496.59	2143.25	1811.12	2266.08	2129.42	1938.96
ENSECAT00000027968	ENSECAG00000025956		1843	1450.34	1882.57	2404.68	1693.75	1609.2	1813.923333
ENSECAT00000014200	ENSECAG00000012593	ATP2A1	1784.7	1648.53	1738.31	1993.75	1799.28	1911.22	1812.631667
ENSECAT00000017254	ENSECAG00000016459		1696.51	1457.57	1826.22	1718.31	2102.87	1994.04	1799.253333
ENSECAT00000025638	ENSECAG00000023911	NDUFB1	2183.35	1827.5	1772.72	1759.18	1400.04	1849.71	1798.75
ENSECAT00000011975	ENSECAG00000011578	COX6A2	1910.57	1726.88	1613.29	1617.13	1935.5	1879.19	1780.426667
ENSECAT00000026923	ENSECAG00000024946		1832.86	1355.29	1790.5	2018.99	1919.87	1722.79	1773.383333
ENSECAT00000029866	ENSECAG00000027700		1575.12	2224.32	1942.8	2005.7	1209.28	1639.48	1766.116667
ENSECAT00000014809	ENSECAG00000014226		3467.14	1458.49	1264.84	966.922	1924.2	1508.12	1764.952
ENSECAT00000010069	ENSECAG00000009780	UQCRB	1775.3	1643.81	1822.96	1813.17	1359.08	1861.63	1712.658333
ENSECAT00000008703	ENSECAG00000007644	PKM	1832.31	1339.65	1801.36	1547.5	1874.17	1665.69	1676.78
ENSECAT00000021693	ENSECAG00000020452	TNNC1	2082.55	1701.9	858.016	1206.18	2019.93	2094.87	1660.574333
ENSECAT00000029859	ENSECAG00000027685		1371.74	1810.98	1687.1	2322.9	881.79	1863.96	1656.411667
ENSECAT00000021383	ENSECAG00000019807	CASQ1	1604.72	1301.99	2276.03	1402.55	1520.16	1753.81	1643.21
ENSECAT00000029502	ENSECAG00000027336		4036.96	745.35	1196.74	630.506	569.469	2537.39	1619.4025
ENSECAT00000003503	ENSECAG00000003667		1652.15	1522.94	1570.88	1249.09	1710.46	2005.79	1618.551667
ENSECAT00000014439	ENSECAG00000013832	DES	1795.21	1759.38	1531.32	1496.13	1540.27	1557.64	1613.325
ENSECAT00000011798	ENSECAG00000011458	COX7A1	1557.78	1653.88	1502.78	1692.57	1684.67	1587.94	1613.27
ENSECAT00000000951	ENSECAG00000001140	ANAPC11	1763.84	1254.93	1634.58	1373.74	1963.93	1655.73	1607.791667
ENSECAT00000014463	ENSECAG00000012593	ATP2A1	1591.59	1403.86	1698.58	1859.63	1510.39	1549.32	1602.228333
ENSECAT00000022345	ENSECAG00000020944	SLC25A4	1756.73	1681.61	1437.9	1723.52	1454.59	1491.85	1591.033333
ENSECAT00000000826	ENSECAG00000001012	LINC00116	1707.27	1179.83	1386.87	1440.12	1948.43	1697.24	1559.96
ENSECAT00000006049	ENSECAG00000006065		1720.92	1240.77	1518.55	1075.22	1932.78	1846.45	1555.781667

ENSECAT00000021149	ENSECAG00000018961	ACTN3	1259.61	1293.53	1635.18	1500.85	1537.41	1650.92	1479.583333
ENSECAT00000023205	ENSECAG00000021705	RPL30	1448.25	1156.41	1441.98	1323.94	1702.51	1621.11	1449.033333
ENSECAT00000016617	ENSECAG00000015659	KLHL41	1533.97	1436.63	1344.06	1407.93	1742.12	1189.97	1442.446667
ENSECAT00000014868	ENSECAG00000014009	TPM2	1613.77	1327.67	2099.69	1048.5	1156.18	1394.52	1440.055
ENSECAT00000017586	ENSECAG00000016758		1359.14	1071.36	1457.03	1475.9	1632.77	1631.09	1437.881667
ENSECAT00000008496	ENSECAG00000008413		1292.87	1348.47	1479.68	1715.94	1261.41	1516.64	1435.835
ENSECAT00000027169	ENSECAG00000025162		1482.46	1297.1	1370.39	1339.78	1468.11	1600.28	1426.353333
ENSECAT00000028024	ENSECAG00000026012	eca-mir-133a	1531.1	755.126	2028.28	860.676	1859.15	1459.45	1415.630333
ENSECAT00000019668	ENSECAG00000018554		1244.42	1112.39	1438.84	1355.16	1636.79	1627.17	1402.461667
ENSECAT00000019116	ENSECAG00000017896	PDLIM3	1318.58	1425.27	1332.59	1327.6	1529.72	1215.14	1358.15
ENSECAT00000001627	ENSECAG00000001815	PYURF	1347.94	1115.36	1244.67	1292.57	1424.44	1570.91	1332.648333
ENSECAT00000010020	ENSECAG00000009744	IRF7	1.37589	1.83482	1.88653	1.44641	1.43087	1.12187	1.516065
ENSECAT00000000605	ENSECAG00000000684	KDM8	1.39021	1.53663	1.40723	1.42856	2.02973	1.2963	1.514776667
ENSECAT00000011525	ENSECAG00000011164	ZNF169	1.45355	1.18167	1.60412	1.50786	1.94483	1.39561	1.514606667
ENSECAT00000013923	ENSECAG00000013399	ROM1	1.71618	1.33174	1.3661	1.68564	1.40195	1.58213	1.513956667
ENSECAT00000021470	ENSECAG00000020088	NOSTRIN	1.60529	1.52978	1.58723	1.0348	1.86109	1.46114	1.513221667
ENSECAT00000025030	ENSECAG00000023009	LCP1	0.525158	1.39128	4.22088	0.774418	1.17403	0.991863	1.512938167
ENSECAT00000013257	ENSECAG00000012336	ARHGAP26	2.95591	0.887745	0.896202	0.33255	2.25517	1.74955	1.5128545

Table 2. Output from Cufflinks. Genes expressed in normal equine muscle at time point2. The top 100 genes with the highest FPKM are shown on the table

Ensembl Transcript ID	Ensembl Gene ID	Gene	FPKM 1	FPKM 2	FPKM 3	FPKM 4	FPKM 5	FPKM 6	AVERAGE
ENSECAT00000029843	ENSECAG00000027698		838375	2584520	2382880	391636	759214	170391	1187836
ENSECAT00000005510	ENSECAG00000005606		339332	434425	462343	147187	174541	181643	289911.8333
ENSECAT00000029864	ENSECAG00000027702		187761	251515	281823	37491.1	98573.2	55033.9	152032.8667
ENSECAT00000029690	ENSECAG00000027524		97574.7	181213	73636.5	46006.1	17070	16332	71972.05
ENSECAT00000027340	ENSECAG00000025328		88852.3	165904	63447.7	64706.1	22608.5	21871.7	71231.71667
ENSECAT00000005564	ENSECAG00000005658		63235.5	60620.7	82299.1	60738.3	102381	49312.3	69764.48333
ENSECAT00000029850	ENSECAG00000027674		48493.7	91734.2	105821	30058	43268.7	11391.2	55127.8
ENSECAT00000029222	ENSECAG00000027056		77781	140335	46583.9	31294.8	12591.7	10986.9	53262.21667
ENSECAT00000029836	ENSECAG00000027670		56608.7	79613.7	93799.1	22159.1	21384.1	9175.59	47123.38167
ENSECAT00000021003	ENSECAG00000019728	TNNC2	37092.1	51086.4	49778.3	32872.4	45924.7	44399.7	43525.6
ENSECAT00000017356	ENSECAG00000016536		24663.8	63983.1	35196.9	72728.7	30340.7	21234.4	41357.93333
ENSECAT00000029275	ENSECAG00000027109		43422.3	83706	28455.5	29575.4	12013.9	10305.8	34579.81667
ENSECAT00000029277	ENSECAG00000027111		43422.3	83706	28455.5	29575.4	12013.9	10305.8	34579.81667
ENSECAT00000029790	ENSECAG00000027624		46396.1	86359.3	27328.6	21666.4	8487.53	7940.21	33029.69
ENSECAT00000018609	ENSECAG00000017437	MYLPF	24310.5	33778.8	36700.9	27848.2	28519.7	30170.6	30221.45
ENSECAT00000029847	ENSECAG00000027680		32786.4	51894.1	54893.1	11315.5	15048.2	7194.05	28855.225
ENSECAT00000029852	ENSECAG00000027682		55951.9	27342.6	12793.3	21707.2	28439.3	14739	26828.88333
ENSECAT00000029837	ENSECAG00000027692	MT-CO2	26542.5	43167.5	40631	30205.8	0	20418.9	26827.61667
ENSECAT00000029490	ENSECAG00000027324		32532.6	58656.2	23056.9	22633.1	7549.44	6302.63	25121.81167
ENSECAT00000029560	ENSECAG00000027394	RN28S	32532.6	58656.2	23056.9	22633.1	7549.44	6302.63	25121.81167
ENSECAT00000016196	ENSECAG00000015494	TNNI2	21022.5	28266.9	27607.5	19418.6	25958.1	24726.9	24500.08333

ENSECAT00000029271	ENSECAG00000027105		31713.9	59955.7	25209.8	16653.7	6549.68	5827.87	24318.44167
ENSECAT00000006002	ENSECAG00000005145	MYL1	20963.1	24028.8	25349.5	20933.7	20058.1	24341.9	22612.51667
ENSECAT00000029848	ENSECAG00000027667		34304.4	23010.5	41438.9	18471	10387.2	6783.66	22399.27667
ENSECAT00000025982	ENSECAG00000023971	TNNT3	19219.1	24862.5	26223.7	17351.5	22461.8	22531.4	22108.33333
ENSECAT00000029835	ENSECAG00000027695	ND3	15836.3	25164.2	19625.6	13241.7	30426.7	16099.1	20065.6
ENSECAT00000029839	ENSECAG00000027668		27782	36903.7	37072.5	7074.26	6910.41	3457.76	19866.77167
ENSECAT00000029856	ENSECAG00000027684	MT-ND1	11253.7	13458.8	15011.7	10501.8	22557.5	11514.6	14049.68333
ENSECAT00000029713	ENSECAG00000027547		18513.5	32371.3	12237.7	12250.7	4581.73	3628.7	13930.605
ENSECAT00000029842	ENSECAG00000027673		18701.7	25361.8	23599.4	3660.38	6317.37	2184.67	13304.22
ENSECAT00000029845	ENSECAG00000027669	MT-CYB	11141.1	12651.9	16875.4	8460.15	16180.6	11157.1	12744.375
ENSECAT00000029854	ENSECAG00000027690		14399.9	24995.9	24285.1	2031.91	5712.95	2654.33	12346.68167
ENSECAT00000023721	ENSECAG00000022051	G3PDH	12502.5	12059.7	13563	6239.65	10848.8	12775.4	11331.50833
ENSECAT00000029834	ENSECAG00000027681	ND2	9374.94	9628.37	11890.3	8526.48	17778	8753.16	10991.875
ENSECAT00000022265	ENSECAG00000020981		9333.95	14929.2	12454.7	8410.32	10292.3	8478.57	10649.84
ENSECAT00000029858	ENSECAG00000027675	MT-ND4	9170.18	11075	14595.9	6268.17	14649.4	7665.08	10570.62167
ENSECAT00000001085	ENSECAG00000001272		9910.92	8544.74	10065	15013.7	8494.81	9917.25	10324.40333
ENSECAT00000004791	ENSECAG00000003121	ALDOA	18573	20465.4	0	19734.4	0	0	9795.466667
ENSECAT00000029844	ENSECAG00000027687		18652.2	13781.7	13864.6	5169.46	4431.57	2627.2	9754.455
ENSECAT00000018976	ENSECAG00000017982	MB	8423.02	9643.88	8889.38	10062.3	13537.2	6891.6	9574.563333
ENSECAT00000015761	ENSECAG00000014949	TPM1	8869.17	10898	12555.9	7809.85	7755.86	9472.92	9560.283333
ENSECAT00000006031	ENSECAG00000006099		9489.52	8198.66	9674.25	9850.77	8191.69	9545.89	9158.463333
ENSECAT00000023921	ENSECAG00000022433	CKM	18235.7	19063.9	0	15710.2	0	0	8834.966667
ENSECAT00000029841	ENSECAG00000027677	ND6	13166.9	15693.6	0	9553.97	0	12315.6	8455.011667
ENSECAT00000000176	ENSECAG00000000207	ACTA1	25623	22254.8	0	0	0	0	7979.633333
ENSECAT00000029218	ENSECAG00000027052	RN18S	8980.4	15648.5	9803.14	5197.37	3424.9	3509.25	7760.593333
ENSECAT00000029223	ENSECAG00000027057	RN18S	8980.4	15648.5	9803.14	5197.37	3424.9	3509.25	7760.593333
ENSECAT00000029516	ENSECAG00000027350	RN18S	8980.4	15648.5	9803.14	5197.37	3424.9	3509.25	7760.593333
ENSECAT00000029666	ENSECAG00000027500	RN18S	8980.4	15648.5	9803.14	5197.37	3424.9	3509.25	7760.593333
ENSECAT00000029394	ENSECAG00000027228	RN18S	8980.4	15643.1	9803.14	5194.77	3424.9	3505.84	7758.691667
ENSECAT00000029403	ENSECAG00000027237	RN18S	8980.4	15643.1	9803.14	5194.77	3424.9	3505.84	7758.691667
ENSECAT00000019534	ENSECAG00000018348	TPT1	6688.92	8722.31	6632.82	7850.17	7674.2	7492.99	7510.235
ENSECAT00000021682	ENSECAG00000020400	ENO3	6337.94	6956.83	7156.21	6798.12	5959.61	7209.16	6736.311667
ENSECAT00000024587	ENSECAG00000022693	SEPT7	54.9845	41.0978	53.5194	65.9318	57.7179	62.0795	6084.190129
ENSECAT00000009054	ENSECAG00000008909		49.9136	51.1791	60.9323	52.7987	55.8032	54.3081	6083.847857
ENSECAT00000006999	ENSECAG00000005768	SEPT11	38.7124	21.5874	21.6766	44.7997	34.7171	37.6061	6065.2999
ENSECAT00000018834	ENSECAG00000017553	SEPT2	19.1952	12.5367	17.372	20.0084	20.9431	23.7179	6051.824757
ENSECAT00000015542	ENSECAG00000014530	SEPT10	13.0859	10.7813	13.1585	15.5888	12.6646	13.3536	6047.947529
ENSECAT00000000444	ENSECAG00000000500		17.0234	29.8501	35.821	19.2781	27.9404	19.4122	6045.760743
ENSECAT00000021867	ENSECAG00000020246	SEPT9	5.85063	11.3399	9.11733	4.46602	8.85578	7.14005	6043.252816
ENSECAT00000021975	ENSECAG00000020248	SEPT4	5.51205	5.83304	4.99521	4.75862	5.44671	5.73801	6040.469091
ENSECAT00000024167	ENSECAG00000022349	SEPT8	2.40531	1.73687	2.7068	2.1427	2.75949	2.46312	6038.459184
ENSECAT00000025351	ENSECAG00000023315	SEPT6	1.16857	4.32898	2.65724	1.25225	1.97661	1.10774	6037.927341
ENSECAT00000022015	ENSECAG00000020718	SEPT5	2.32862	1.59444	2.31954	2.12618	1.81738	1.34341	6037.647081

ENSECAT00000018782	ENSECAG00000017362	SEPT14	0	0	0	0	0.047197 6	0.032297 4	6037.297071
ENSECAT00000009959	ENSECAG00000009390	SEPT1	0.454846	5.76598	3.37626	0.679292	0.764812	0.348911	6037.055729
ENSECAT00000020118	ENSECAG00000018692	SEPT12	0	0	0	0	0	0.031281 5	6037.004469
ENSECAT00000025376	ENSECAG00000023601	SEPT3	0.021057 8	0.178791	0.025009 4	0	0.012597 6	0	6035.748208
ENSECAT00000019881	ENSECAG00000018637	MARC1	17.8993	13.7024	15.7386	19.4781	20.4976	23.6037	6024.988529
ENSECAT00000027025	ENSECAG00000025028	MARCH2	10.9743	14.9544	16.4024	12.3967	14.1696	12.742	6020.948486
ENSECAT00000016745	ENSECAG00000015805	MARCH7	11.4633	8.20702	8.91466	13.2954	12.472	12.582	6019.562054
ENSECAT00000029139	ENSECAG00000012069	MARCH6	12.6045	3.77882	6.83812	11.5967	12.1766	12.9958	6018.42722
ENSECAT00000012487	ENSECAG00000012069	MARCH6	10.5893	6.66165	7.54756	10.119	9.0641	11.3188	6017.757201
ENSECAT00000027143	ENSECAG00000025138	MARCH9	2.4495	3.11961	3.25463	3.16198	2.52569	4.11733	6012.946963
ENSECAT00000002899	ENSECAG00000002745	MARCH8	2.23935	1.5674	2.27351	2.22577	1.85582	2.39958	6011.937347
ENSECAT00000007639	ENSECAG00000007518	MARCH3	1.58155	3.11764	1.7662	1.2742	1.75069	1.70283	6011.027587
ENSECAT00000026799	ENSECAG00000024858	MARCH1 1	0	0	0	0	0	0	6010.571429
ENSECAT00000000932	ENSECAG00000000705	MARCH1 0	0	0	0	0	0.008407 13	0.017978 8	6010.432341
ENSECAT00000000881	ENSECAG00000000705	MARCH1 0	0	0	0	0	0	0	6010.428571
ENSECAT00000007391	ENSECAG00000007360	MARCH4	0	0	0	0	0.01565	0	6009.573664
ENSECAT00000015677	ENSECAG00000015012	MARCH1	0	0.096762 7	0.21318	0	0.091800 4	0.104694	6009.215205
ENSECAT00000029866	ENSECAG00000027700		7826.09	8982.69	8055.42	2478.92	3681.25	2411.53	5572.65
ENSECAT00000029015	ENSECAG00000026987	NDUFA4	4336.62	6750.73	5957.95	5074.56	6504.6	4649.46	5545.653333
ENSECAT00000017402	ENSECAG00000016604		0	182.771	46.0192	15931.6	10378.7	5076.76	5269.308367
ENSECAT00000022404	ENSECAG00000020669	LDHA	4591.98	5009.55	5302.53	5937.67	4961.99	5137.04	5156.793333
ENSECAT00000009354	ENSECAG00000009190	TMSB4	3689.48	7454.31	5949.19	4560.83	4940.03	2652.46	4874.383333
ENSECAT00000029859	ENSECAG00000027685		6782.61	7752.19	7745.59	1692.92	2696.46	1980.2	4774.995
ENSECAT00000005912	ENSECAG00000005539	PGAM2	3919.87	5744.4	5140.57	4186.8	4436.95	4796.95	4704.256667
ENSECAT00000018208	ENSECAG00000017292	ATP5I	3228.02	5588.58	5359.48	3824.02	5032.25	3750.76	4463.851667
ENSECAT00000029862	ENSECAG00000027694	MT-ND5	7839.19	5537.98	0	5062.02	0	7395.5	4305.781667
ENSECAT00000008746	ENSECAG00000008542		3528.17	4639.34	3942.44	2257.44	4614.13	4886.17	3977.948333
ENSECAT00000029838	ENSECAG00000027678		5575.01	5147.49	5705.5	1346.74	2703.56	1607.78	3681.013333
ENSECAT00000029857	ENSECAG00000027683		8142.99	4970.7	4456.28	1902.04	1449.37	691.552	3602.155333
ENSECAT00000007630	ENSECAG00000007612	PPP1R1A	2919.15	3926.06	4137.52	3065.22	3693.46	3778.35	3586.626667
ENSECAT00000007368	ENSECAG00000007378	C14orf2	2817.74	4291.24	3583.35	3083.27	4191.33	2969.21	3489.356667
ENSECAT00000016514	ENSECAG00000015581	TPI1	3077.65	3682.05	3764.76	3076.15	3050.9	3763.22	3402.455
ENSECAT00000006892	ENSECAG00000006913		4044.55	3724.2	3148.91	1789.22	3844.29	3537.53	3348.116667
ENSECAT00000000003	ENSECAG00000000008	RPLP2	2661.11	4407.01	4221.01	2797.78	3138.25	2805.75	3338.485
ENSECAT00000029543	ENSECAG00000027377	5_8S_rR NA	3437.48	7343.98	5622.85	1970.37	459.119	1136.82	3328.4365
ENSECAT00000003503	ENSECAG00000003667		3310.46	4433.62	3440.24	1768.71	2705.04	2275.75	2988.97

Table 3. Output from Cufflinks. Genes expressed in normal equine muscle at time point3. The top 100 genes with the highest FPKM are shown on the table

Ensembl Transcript ID	Ensembl Gene ID	Gene	FPKM 1	FPKM 2	FPKM 3	FPKM 4	FPKM 5	FPKM 6	AVERAGE
ENSECAT00000029843	ENSECAG00000027698		484483	1.19E+06	1.19E+06	2.76E+06	685890	558458	1143296.833
ENSECAT00000005510	ENSECAG00000005606		149819	388856	396154	412155	372891	144997	310812
ENSECAT00000029864	ENSECAG00000027702		67393.8	248770	182347	375107	168342	83376.2	187556
ENSECAT00000027340	ENSECAG00000025328		26021.5	56152.2	75271.5	119930	73128.4	222558	95510.26667
ENSECAT00000029690	ENSECAG00000027524		17620.6	60212.5	87213.6	137605	77341	170419	91735.28333
ENSECAT00000005564	ENSECAG00000005658		51501.7	90352.2	68741.4	122128	65889.9	46974.1	74264.55
ENSECAT00000029222	ENSECAG00000027056		18527.1	38467.2	61753.5	90563.9	53165.9	180552	73838.26667
ENSECAT00000029850	ENSECAG00000027674		27284.4	58062.9	69562.7	122228	30313.4	37265.3	57452.78333
ENSECAT00000029790	ENSECAG00000027624		12581.9	24295	41284.7	59975.1	36475.2	164710	56553.65
ENSECAT00000029275	ENSECAG00000027109		13642.7	23316	34937.1	52830.4	28471.4	134352	47924.93333
ENSECAT00000029277	ENSECAG00000027111		13642.7	23316	34937.1	52830.4	28471.4	134352	47924.93333
ENSECAT00000029836	ENSECAG00000027670		19050.5	78075.8	50804.9	77879.6	25732	27420.4	46493.86667
ENSECAT00000021003	ENSECAG00000019728	TNNC2	44578.3	45989.8	44163.2	48421.8	43858.5	42999.3	45001.81667
ENSECAT00000017356	ENSECAG00000016536		51590.6	33350.1	64409	35523.5	26157.5	17835.9	38144.43333
ENSECAT00000029852	ENSECAG00000027682		11543.2	68293.8	45642.6	57332.3	35300.3	4632.83	37124.17167
ENSECAT00000029490	ENSECAG00000027324		8613.57	18183.1	27215.9	46383	23304.7	79233.3	33822.26167
ENSECAT00000029560	ENSECAG00000027394	RN28S	8613.57	18183.1	27215.9	46383	23304.7	79233.3	33822.26167
ENSECAT00000018609	ENSECAG00000017437	MYLPF	29316.8	29799.2	28938.9	32208.6	29116.8	28690.5	29678.46667
ENSECAT00000029271	ENSECAG00000027105		6139.69	19299.2	27350.5	44556.3	22618.8	57946	29651.74833
ENSECAT00000029837	ENSECAG00000027692	MT-CO2	39135.6	36206.1	29585.5	0	20725.7	41016.4	27778.21667
ENSECAT00000029847	ENSECAG00000027680		14040	49745.3	31316	40646	13878.8	16114.2	27623.38333
ENSECAT00000029848	ENSECAG00000027667		9115.57	47961.9	41161.7	30942.1	18720.7	9002.48	26150.74167
ENSECAT00000016196	ENSECAG00000015494	TNNI2	23760	25148.6	27436	28398.9	24307.2	25069	25686.61667
ENSECAT00000029543	ENSECAG00000027377	5_8S_rRNA	2346.56	2508.21	5488.6	12724.2	6021.67	110750	23306.54
ENSECAT00000025982	ENSECAG00000023971	TNNT3	21860.2	21703.1	22496.9	24166.3	22035	19553.5	21969.16667
ENSECAT00000029835	ENSECAG00000027695	ND3	22496.4	24590.8	17083.6	30261.2	16015.5	19652.4	21683.31667
ENSECAT00000006002	ENSECAG00000005145	MYL1	21064.4	22664.1	22284.6	21560.8	22537.2	19643.2	21625.71667
ENSECAT00000029839	ENSECAG00000027668		7621.23	33356.2	24638.6	31135.7	10971.5	9670.34	19565.595
ENSECAT00000029713	ENSECAG00000027547		4986.81	10220.7	15229.7	25243.5	13238.8	44927.9	18974.56833
ENSECAT00000029842	ENSECAG00000027673		2510.7	19745.7	23005.4	37858.3	15113.6	3922.64	17026.05667
ENSECAT00000029394	ENSECAG00000027228	RN18S	4027	5938.88	9669.82	22597	9853.43	49508.5	16932.43833
ENSECAT00000029403	ENSECAG00000027237	RN18S	4027	5938.88	9669.82	22597	9853.43	49508.5	16932.43833
ENSECAT00000029218	ENSECAG00000027052	RN18S	4027	5938.88	9669.82	22589.4	9847.1	49500.9	16928.85
ENSECAT00000029223	ENSECAG00000027057	RN18S	4027	5938.88	9669.82	22589.4	9847.1	49500.9	16928.85
ENSECAT00000029516	ENSECAG00000027350	RN18S	4027	5938.88	9669.82	22589.4	9847.1	49500.9	16928.85
ENSECAT00000029666	ENSECAG00000027500	RN18S	4027	5938.88	9669.82	22589.4	9847.1	49500.9	16928.85
ENSECAT00000029856	ENSECAG00000027684	MT-ND1	13474	16486.3	11874	22553.6	11559.1	13919.1	14977.68333

ENSECAT00000029845	ENSECAG00000027669	MT-CYB	12942.9	15218.2	10720.4	16691.6	11257.4	14058.3	13481.46667
ENSECAT00000023721	ENSECAG00000022051	G3PDH	10421.6	13110	14510.7	11706.6	14393.2	9646.08	12298.03
ENSECAT00000022265	ENSECAG00000020981		9929.63	14381.1	12795.5	15055.6	9640.69	10936.6	12123.18667
ENSECAT00000029834	ENSECAG00000027681	ND2	10347.1	13004.1	9611.16	17724.4	9073.07	9877.05	11606.14667
ENSECAT00000029854	ENSECAG00000027690		5141.18	18001.3	15004.5	19974.9	5665.63	4102.58	11315.015
ENSECAT00000029844	ENSECAG00000027687		3072.66	19610.7	13785.3	18093.2	9179.66	3799.21	11256.78833
ENSECAT00000029502	ENSECAG00000027336		1038	2003.51	4226.28	11632.3	7498.27	40138.5	11089.47667
ENSECAT00000029858	ENSECAG00000027675	MT-ND4	10769.8	10984.4	8872.32	14944.8	7936.34	12585.9	11015.59333
ENSECAT00000018976	ENSECAG00000017982	MB	11026.4	12744.5	10051.8	13908.3	7258.14	10632.2	10936.89
ENSECAT00000029841	ENSECAG00000027677	ND6	17227.1	0	12754.3	0	13215.1	18356	10258.75
ENSECAT00000001085	ENSECAG00000001272		7366.57	10027.3	11141.5	8840.88	11399.9	6727.31	9250.576667
ENSECAT00000015761	ENSECAG00000014949	TPM1	9499.11	9532.86	9579.74	7735.95	9610.75	9286.08	9207.415
ENSECAT00000006031	ENSECAG00000006099		7089.84	9600.44	10700.2	8497.72	10914.6	6486.45	8881.541667
ENSECAT00000000176	ENSECAG00000000207	ACTA1	24936.5	0	0	0	0	23299	8039.25
ENSECAT00000019534	ENSECAG00000018348	TPT1	8002.37	8805.87	8648.49	8576.98	6493.41	6101.21	7771.388333
ENSECAT00000029866	ENSECAG00000027700		2458.13	10503.1	8607.71	11362.9	6216.15	2642.93	6965.153333
ENSECAT00000021682	ENSECAG00000020400	ENO3	5969.48	6950.12	7925.55	6259.53	7771.94	5793.83	6778.408333
ENSECAT00000029015	ENSECAG00000026987	NDUFA4	5665.89	5877.76	5798.27	7133.28	4902.44	6645.94	6003.93
ENSECAT00000004791	ENSECAG00000003121	ALDOA	18507.4	0	0	0	0	15451.5	5659.816667
ENSECAT00000023921	ENSECAG00000022433	CKM	18843.2	0	0	0	0	14876.9	5620.016667
ENSECAT00000029859	ENSECAG00000027685		1604.61	6757.26	6795.56	9614.77	5493.34	2642.93	5484.745
ENSECAT00000029862	ENSECAG00000027694	MT-ND5	8064.74	0	6615.01	0	7873	7305.18	4976.321667
ENSECAT00000018208	ENSECAG00000017292	ATP5I	4106.01	5099.31	4820.16	6864.73	3580.44	5287.8	4959.741667
ENSECAT00000022404	ENSECAG00000020669	LDHA	5004.9	4494.05	6191.25	5078.77	5128.04	3437.07	4889.013333
ENSECAT00000005912	ENSECAG00000005539	PGAM2	4828.72	5027.05	5429.44	4878.34	4671.81	4350.73	4864.348333
ENSECAT00000029838	ENSECAG00000027678		1469.99	6345.43	4767.33	8107.75	5192.83	2094.17	4662.916667
ENSECAT00000008746	ENSECAG00000008542		3827.77	5221.32	4999.68	4978.09	4769.86	3774.3	4595.17
ENSECAT00000029857	ENSECAG00000027683		1606.44	8811.08	6021.01	5028.8	2742.87	1581.6	4298.633333
ENSECAT00000009354	ENSECAG00000009190	TMSB4	3483.61	4617.26	4306.75	5757.83	3419.36	4081.2	4277.668333
ENSECAT00000007368	ENSECAG00000007378	C14orf2	3453.1	4567.29	3528.54	5211.28	3051.33	3944.15	3959.281667
ENSECAT00000006892	ENSECAG00000006913		3449.83	5331.92	3724.33	3784.53	3914.98	3117.98	3887.261667
ENSECAT00000007630	ENSECAG00000007612	PPP1R1A	3409.73	3833.75	4108.74	4023.77	3408.41	3646.98	3738.563333
ENSECAT00000000003	ENSECAG00000000008	RPLP2	3511.08	4401.8	3412.37	3742.9	2823.15	3741.92	3605.536667
ENSECAT00000003503	ENSECAG00000003667		1853.76	4577.13	4751.8	5090.46	3619.87	1410.46	3550.58
ENSECAT00000014809	ENSECAG00000014226		1199.61	1263.2	1920.05	3225.47	1692.49	11055.6	3392.736667
ENSECAT00000016514	ENSECAG00000015581	TPI1	3082.17	3442.82	3464.6	3363.21	3535.1	2981.08	3311.496667
ENSECAT00000026040	ENSECAG00000024000	PYGM	2942.44	2801.71	3086.27	2594.79	3486.38	2467.25	2896.473333
ENSECAT00000027884	ENSECAG00000025872	eca-mir-1842	418.45	3510.34	3479.8	3916.37	4857.94	411.244	2765.690667
ENSECAT00000027983	ENSECAG00000025971	eca-mir-1842	418.45	3510.34	3479.8	3916.37	4857.94	411.244	2765.690667
ENSECAT00000014627	ENSECAG00000013777	YBX3	2512.71	2869.46	2824.62	2774.47	2649.77	2384.85	2669.313333
ENSECAT00000014868	ENSECAG00000014009	TPM2	2314.92	2768.98	2752.09	3058.89	2088.54	2627.69	2601.851667
ENSECAT00000029865	ENSECAG00000027693		535.54	4054.34	2715.11	4245.86	2792.92	820.517	2527.381167
ENSECAT00000029860	ENSECAG00000027696		1195.49	3866.09	3644.3	3423.86	2211.99	813.393	2525.853833

ENSECAT00000006049	ENSECAG00000006065		2036.8	2840.85	2698.95	2693.07	2677.09	1854.73	2466.915
ENSECAT00000008314	ENSECAG00000007867	MYL2	1961.32	3439.48	3045.52	1916.25	1960.29	2473.93	2466.131667
ENSECAT00000029375	ENSECAG00000027209		128.997	486.793	803.989	2525.18	1615.27	9177.74	2456.328167
ENSECAT00000000951	ENSECAG00000001140	ANAPC11	1763.06	2915.71	2523.73	2978.52	2748.74	1670.97	2433.455
ENSECAT00000021693	ENSECAG00000020452	TNNC1	1856.72	3667.07	2753.19	1781.48	2156.69	2222.16	2406.218333
ENSECAT00000017584	ENSECAG00000016730		1523.8	3105.32	1476.92	4037.41	2554.51	1651.1	2391.51
ENSECAT00000008703	ENSECAG00000007644	PKM	2078.99	2375.55	2821.86	2156.33	2820.02	1978.61	2371.893333
ENSECAT00000026923	ENSECAG00000024946		2547.59	2425.11	2763.42	1872.21	2015.19	2302.17	2320.948333
ENSECAT00000017694	ENSECAG00000016739	MYOZ1	2307.38	2461.26	2435.34	2365.87	2259.64	1998.45	2304.656667
ENSECAT00000025638	ENSECAG00000023911	NDUFB1	1852.69	2356.35	2589.29	2982.56	1361.02	2037.25	2196.526667
ENSECAT00000011975	ENSECAG00000011578	COX6A2	1833.36	2216.77	2056.24	2495.75	1730.61	2254.07	2097.8
ENSECAT00000021383	ENSECAG00000019807	CASQ1	2096.59	2005.16	2309.92	1921.36	2328.48	1708.43	2061.656667
ENSECAT00000022182	ENSECAG00000020407	MYBPC2	1655.65	2078.37	2589.56	1557.47	2652	1751.67	2047.453333
ENSECAT00000028889	ENSECAG00000026905	RPL37	1684.26	2117.46	2145.75	2276.18	2011.62	1791.43	2004.45
ENSECAT00000011798	ENSECAG00000011458	COX7A1	1738.64	2149.15	1530.75	2747.35	1435.26	2384.41	1997.593333
ENSECAT00000028024	ENSECAG00000026012	eca-mir-133a	1131.2	1926.33	2372.89	2269.03	3056.34	1026.2	1963.665
ENSECAT00000027968	ENSECAG00000025956		1108.29	2427.26	2327.86	2082.62	2401.86	1331.05	1946.49
ENSECAT00000000826	ENSECAG00000001012	LINC00116	1856.86	2169.66	2111.11	1766.19	1963.07	1707.59	1929.08
ENSECAT00000029840	ENSECAG00000027699		1602.07	1748.13	1070.22	2124.09	668.533	3954.64	1861.2805
ENSECAT00000027169	ENSECAG00000025162		1730.1	2011.13	1808.91	2054.79	1496.36	1886.29	1831.263333

Table 4. Comparisons performed between time point t1 and t2. Training effect within group. The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl GENE ID	GENE Name	logFC	PValue	FDR
Case t1 X Case t2	ENSECAG00000011484	EPS8L2	-4.78128379	1.23738E-52	1.4272E-48
Case t1 X Case t2	ENSECAG00000007809	OR2T27	-3.143150421	5.35865E-30	3.09033E-26
Case t1 X Case t2	ENSECAG00000016846	ENSECAG00000016846	-2.598850406	2.99273E-22	1.15061E-18
Case t1 X Case t2	ENSECAG00000020221	KNDC1	-3.2791048	6.42218E-22	1.85184E-18
Case t1 X Case t2	ENSECAG00000026913	ENSECAG00000026913	-2.10756729	1.96691E-18	4.53727E-15
Case t1 X Case t2	ENSECAG00000024589	ENSECAG00000024589	-2.311686808	3.89025E-17	7.47837E-14
Case t1 X Case t2	ENSECAG00000013836	PKLR	2.298607192	4.91875E-16	8.1047E-13
Case t1 X Case t2	ENSECAG00000019552	SLCO1C1	2.33545454	3.87642E-15	5.58883E-12
Case t1 X Case t2	ENSECAG00000013781	KIAA1211	-1.697206356	3.77175E-14	4.83371E-11
Case t1 X Case t2	ENSECAG00000006267	NR1D1	-1.346418978	7.26095E-11	7.61343E-08
Case t1 X Case t2	ENSECAG00000019557	LDLR	-1.388341109	6.80243E-11	7.61343E-08
Case t1 X Case t2	ENSECAG00000014500	ART5	1.271377618	2.75373E-10	2.64679E-07
Case t1 X Case t2	ENSECAG00000021096	NRXN2	-1.564598167	4.40512E-10	3.90836E-07
Case t1 X Case t2	ENSECAG00000024790	ENSECAG00000024790	1.322693497	2.19685E-09	1.80989E-06
Case t1 X Case t2	ENSECAG00000020485	PCOLCE2	1.296580385	2.54646E-09	1.95806E-06
Case t1 X Case t2	ENSECAG00000027324	RNA28S1	1.36718335	4.25282E-09	3.06575E-06
Case t1 X Case t2	ENSECAG00000023949	CCL2	1.515983122	1.90284E-08	1.29102E-05
Case t1 X Case t2	ENSECAG00000016634	RPS2	-1.170961299	9.40957E-08	6.02944E-05
Case t1 X Case t2	ENSECAG00000008710	FCER1A	1.588999592	1.08387E-07	6.57966E-05
Case t1 X Case t2	ENSECAG00000007161	CYP2F1	1.294710352	1.40488E-07	8.10194E-05
Case t1 X Case t2	ENSECAG00000002962	ADIPOQ	1.238907368	1.63498E-07	8.97993E-05
Case t1 X Case t2	ENSECAG00000014226	ENSECAG00000014226	1.062158811	2.28842E-07	0.000119976
Case t1 X Case t2	ENSECAG00000016225	TMEM144	1.665037785	2.63014E-07	0.000126209
Case t1 X Case t2	ENSECAG00000025328	RNA45S5	1.056872193	2.67001E-07	0.000126209
Case t1 X Case t2	ENSECAG00000027524	ENSECAG00000027524	1.058301003	2.73558E-07	0.000126209
Case t1 X Case t2	ENSECAG00000013022	ACTC1	1.097707766	3.73807E-07	0.000165827
Case t1 X Case t2	ENSECAG00000021019	CPXM1	1.102435473	4.71363E-07	0.000201359
Case t1 X Case t2	ENSECAG00000020878	MNDA	1.104024877	4.97943E-07	0.000205117
Case t1 X Case t2	ENSECAG00000013687	SMG1	-1.110581229	1.55494E-06	0.000618435
Case t1 X Case t2	ENSECAG00000027056	ENSECAG00000027056	0.982814572	1.89724E-06	0.000729427
Case t1 X Case t2	ENSECAG00000016560	CRABP2	1.00124932	2.22157E-06	0.000826569
Case t1 X Case t2	ENSECAG00000019220	PROX1	-1.1209383	3.81444E-06	0.001374867
Case t1 X Case t2	ENSECAG00000020059	FGF4	1.222608401	3.99579E-06	0.001396591
Case t1 X Case t2	ENSECAG00000016782	MYO1F	-0.960312958	4.84248E-06	0.001595806
Case t1 X Case t2	ENSECAG00000017431	MYOZ2	0.916496424	4.74411E-06	0.001595806
Case t1 X Case t2	ENSECAG00000027624	ENSECAG00000027624	0.956400782	5.02439E-06	0.001609758
Case t1 X Case t2	ENSECAG00000008916	ANKRD2	0.942918425	5.80386E-06	0.001805296
Case t1 X Case t2	ENSECAG00000021324	NR6A1	-1.022308695	5.94774E-06	0.001805296
Case t1 X Case t2	ENSECAG00000016582	SSPO	-1.118591246	7.19798E-06	0.002128757

Case t1 X Case t2	ENSECAG00000017807	NAV2	-0.916963886	8.5914E-06	0.002416907
Case t1 X Case t2	ENSECAG00000024660	DPF3	-0.943473031	8.5434E-06	0.002416907
Case t1 X Case t2	ENSECAG00000027699	ENSECAG00000027699	0.869780299	1.27308E-05	0.003496134
Case t1 X Case t2	ENSECAG00000026887	TSPAN8	0.871818503	1.39415E-05	0.003739571
Case t1 X Case t2	ENSECAG00000013517	LCOR	-0.902060461	1.5827E-05	0.004148821
Case t1 X Case t2	ENSECAG00000009126	NOS1	-0.849016463	2.32812E-05	0.005967233
Case t1 X Case t2	ENSECAG00000013728	MFSD2A	1.092417355	2.42829E-05	0.006088682
Case t1 X Case t2	ENSECAG00000015909	MYBPHL	0.965982026	2.52362E-05	0.006193059
Case t1 X Case t2	ENSECAG00000013913	PCK2	1.011936261	3.0714E-05	0.007115326
Case t1 X Case t2	ENSECAG00000018927	ANGPTL4	1.315852475	3.00746E-05	0.007115326
Case t1 X Case t2	ENSECAG00000018973	ENTHD1	0.882669224	3.0845E-05	0.007115326
Case t1 X Case t2	ENSECAG00000011254	SMG1	-0.881040155	3.25071E-05	0.007332647
Case t1 X Case t2	ENSECAG00000019769	PPARGC1B	-0.845040345	3.30586E-05	0.007332647
Case t1 X Case t2	ENSECAG00000008881	LYVE1	1.081690146	4.2435E-05	0.009063802
Case t1 X Case t2	ENSECAG00000023163	ENSECAG00000023163	-0.823719561	4.24087E-05	0.009063802
Case t1 X Case t2	ENSECAG00000009390	SEPT1	0.935311453	4.85474E-05	0.009823606
Case t1 X Case t2	ENSECAG00000023118	MCC	-0.839393141	4.82161E-05	0.009823606
Case t1 X Case t2	ENSECAG00000024143	MYO7A	-0.976939527	4.7703E-05	0.009823606
Case t1 X Case t2	ENSECAG00000001101	SESN1	0.791257883	5.67922E-05	0.01092791
Case t1 X Case t2	ENSECAG00000010266	NFAT5	-0.831615204	5.68471E-05	0.01092791
Case t1 X Case t2	ENSECAG00000019083	SORL1	-0.930650876	5.52259E-05	0.01092791
Case t1 X Case t2	ENSECAG00000000171	NPAS2	-0.988864073	6.25025E-05	0.011818092
Case t1 X Case t2	ENSECAG00000022544	KMT2C	-0.803892872	6.43492E-05	0.011971036
Case t1 X Case t2	ENSECAG00000019111	CD163	0.874990518	7.04406E-05	0.012499409
Case t1 X Case t2	ENSECAG00000023280	FADS1	-0.910205273	6.8483E-05	0.012499409
Case t1 X Case t2	ENSECAG00000024267	TFRC	-0.788703171	6.993E-05	0.012499409
Case t1 X Case t2	ENSECAG00000007663	GIMAP5	0.924855529	7.46506E-05	0.013045752
Case t1 X Case t2	ENSECAG00000018324	SLC43A2	0.794473572	7.76775E-05	0.013372127
Case t1 X Case t2	ENSECAG00000007047	ASPN	0.891294062	7.94172E-05	0.013470556
Case t1 X Case t2	ENSECAG00000007582	GIMAP7	0.915221414	8.58579E-05	0.014146927
Case t1 X Case t2	ENSECAG00000023076	GPCPD1	0.777738691	8.55911E-05	0.014146927
Case t1 X Case t2	ENSECAG00000006043	PHC3	-0.826927649	8.81954E-05	0.014327404
Case t1 X Case t2	ENSECAG00000000040	IKZF2	-0.842827447	9.06308E-05	0.014518552
Case t1 X Case t2	ENSECAG00000017294	SNTB1	-0.771538188	9.34903E-05	0.014771462
Case t1 X Case t2	ENSECAG00000009759	TNRC6B	-0.783049732	0.000100611	0.015681693
Case t1 X Case t2	ENSECAG00000017238	MTC1	0.893145088	0.000104608	0.016087346
Case t1 X Case t2	ENSECAG00000023387	FRRS1	-0.771461128	0.000108041	0.016396608
Case t1 X Case t2	ENSECAG00000020410	FREM2	-0.840703902	0.000113946	0.017068175
Case t1 X Case t2	ENSECAG00000015076	INO80D	-0.841658102	0.00012068	0.017845165
Case t1 X Case t2	ENSECAG00000008221	S100A4	0.777938675	0.000147727	0.021298474
Case t1 X Case t2	ENSECAG00000016173	TK1	0.898468848	0.000147266	0.021298474
Case t1 X Case t2	ENSECAG00000004773	TET2	-0.814472475	0.000154296	0.021971001
Case t1 X Case t2	ENSECAG00000024549	KIAA2018	-0.761033967	0.000161431	0.022706626

Case t1 X Case t2	ENSECAG00000017015	MYH11	-0.766792538	0.000165872	0.023050243
Case t1 X Case t2	ENSECAG00000005666	ZBED6	-0.755511021	0.000173707	0.023851587
Case t1 X Case t2	ENSECAG00000002206	SPRY3	-0.938833481	0.000191771	0.026022189
Case t1 X Case t2	ENSECAG00000012155	GAN	-0.814082401	0.000197192	0.026446642
Case t1 X Case t2	ENSECAG00000023650	ABAT	-0.872861552	0.000204686	0.027136222
Case t1 X Case t2	ENSECAG00000009535	FCER1G	0.801428029	0.000213226	0.027947086
Case t1 X Case t2	ENSECAG00000013723	SLC7A8	-0.744908856	0.000215895	0.027979051
Case t1 X Case t2	ENSECAG00000021790	LNPEP	-0.75209736	0.000222891	0.028564707
Case t1 X Case t2	ENSECAG00000017042	ENSECAG00000017042	0.854906711	0.000226468	0.028704213
Case t1 X Case t2	ENSECAG00000020053	SLC4A4	-0.792815614	0.000230056	0.028831319
Case t1 X Case t2	ENSECAG00000023632	RGS10	0.741984139	0.00023247	0.028831319
Case t1 X Case t2	ENSECAG00000018191	GUCY1A2	-0.818695963	0.000244669	0.030021411
Case t1 X Case t2	ENSECAG00000017946	KMT2D	-0.73014834	0.000258841	0.031425966
Case t1 X Case t2	ENSECAG00000000762	DDI2	-0.754614254	0.000272737	0.032230268
Case t1 X Case t2	ENSECAG00000016683	SMYD1	-0.718670775	0.000273848	0.032230268
Case t1 X Case t2	ENSECAG00000017427	OTOR	0.974489221	0.000270382	0.032230268
Case t1 X Case t2	ENSECAG00000000237	DGKH	-0.819461561	0.0002859	0.033308769
Case t1 X Case t2	ENSECAG00000007668	ENSECAG00000007668	0.925416797	0.000289727	0.033417098

Table 5. Comparisons performed between time point t2 and t3. Exercise effect within group. The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl GENE ID	GENE	logFC	PValue	FDR
Case t2 X Case t3	ENSECAG00000014260	FOS	7.598763208	1.1264E-170	1.2992E-166
Case t2 X Case t3	ENSECAG00000014338	EGR1	7.033589317	7.7632E-161	4.477E-157
Case t2 X Case t3	ENSECAG00000018002	FOSB	8.638430603	2.0631E-146	7.932E-143
Case t2 X Case t3	ENSECAG00000003816	JUNB	4.473073737	4.89699E-81	1.41205E-77
Case t2 X Case t3	ENSECAG00000017203	BTG2	4.275525933	2.09277E-72	4.8276E-69
Case t2 X Case t3	ENSECAG00000014232	NR4A2	4.130895224	5.90984E-57	1.13607E-53
Case t2 X Case t3	ENSECAG00000010560	MGSA	4.774409938	1.12126E-56	1.84752E-53
Case t2 X Case t3	ENSECAG00000011486	ATF3	3.418073922	1.55216E-51	2.23783E-48
Case t2 X Case t3	ENSECAG00000010818	EGR2	3.999060604	4.51143E-50	5.78165E-47
Case t2 X Case t3	ENSECAG00000009012	IER2	3.304921184	6.50951E-49	7.50807E-46
Case t2 X Case t3	ENSECAG00000009958	DUSP1	3.09207054	4.50384E-46	4.72248E-43
Case t2 X Case t3	ENSECAG00000004180	HSPA6	3.755188486	4.94437E-45	4.75237E-42
Case t2 X Case t3	ENSECAG00000024993	CYR61	3.087759541	1.82075E-43	1.61542E-40
Case t2 X Case t3	ENSECAG00000010613	KLF4	2.830283968	1.69338E-37	1.39511E-34
Case t2 X Case t3	ENSECAG00000001249	SOCS3	3.838014794	1.60919E-34	1.23736E-31

Case t2 X Case t3	ENSECAG00000024055	JUN	2.576022754	3.60964E-33	2.6021E-30
Case t2 X Case t3	ENSECAG00000013836	PKLR	-5.039210367	6.11038E-33	4.14571E-30
Case t2 X Case t3	ENSECAG00000019552	SLCO1C1	-5.987065861	3.30454E-32	2.11748E-29
Case t2 X Case t3	ENSECAG00000020971	RND1	3.001285982	6.22398E-32	3.77829E-29
Case t2 X Case t3	ENSECAG00000016611	KLF2	2.367888106	9.14709E-28	5.27513E-25
Case t2 X Case t3	ENSECAG00000007242	DUSP2	4.011343384	3.77467E-27	2.07319E-24
Case t2 X Case t3	ENSECAG00000011488	MIP-2BETA	3.905463879	4.3806E-27	2.29663E-24
Case t2 X Case t3	ENSECAG00000020059	FGF4	-4.04051323	2.29056E-25	1.14867E-22
Case t2 X Case t3	ENSECAG00000005487	IGLC7	4.043096642	6.73026E-24	3.23445E-21
Case t2 X Case t3	ENSECAG00000017436	TREM1	2.788982413	3.46108E-22	1.5968E-19
Case t2 X Case t3	ENSECAG00000011496	NFKBIZ	2.123171438	5.58381E-22	2.47706E-19
Case t2 X Case t3	ENSECAG00000011895	ILT11B	2.629998058	3.49189E-20	1.49168E-17
Case t2 X Case t3	ENSECAG00000013457	CSF3R	2.168765643	1.38393E-18	5.70079E-16
Case t2 X Case t3	ENSECAG00000013081	MMP9	2.386807876	2.43589E-18	9.68813E-16
Case t2 X Case t3	ENSECAG00000004066	APOLD1	2.118307517	4.09667E-18	1.57503E-15
Case t2 X Case t3	ENSECAG00000013781	KIAA1211	-2.520995413	6.05905E-18	2.25436E-15
Case t2 X Case t3	ENSECAG00000013730	ID1	1.818135687	1.84767E-17	6.65969E-15
Case t2 X Case t3	ENSECAG00000023949	CCL2	2.009780927	2.03138E-17	7.09998E-15
Case t2 X Case t3	ENSECAG00000020122	SELL	2.095629557	1.59054E-16	5.39566E-14
Case t2 X Case t3	ENSECAG00000001471	GPR34	2.451200089	2.2118E-16	7.28884E-14
Case t2 X Case t3	ENSECAG00000004312	IL1RN	2.20639216	6.33413E-16	2.02938E-13
Case t2 X Case t3	ENSECAG00000025128	PIWIL2	-2.795650277	8.42879E-16	2.62751E-13
Case t2 X Case t3	ENSECAG00000023741	EGR3	1.829703695	3.47678E-15	1.0553E-12
Case t2 X Case t3	ENSECAG00000026998	LRRC10B	-2.702773129	8.86041E-15	2.62041E-12
Case t2 X Case t3	ENSECAG00000007258	IGHG1	2.507466372	1.56053E-14	4.49979E-12
Case t2 X Case t3	ENSECAG00000003072	C8orf4	1.640009076	3.07419E-14	8.64822E-12
Case t2 X Case t3	ENSECAG00000009625	IGHA1	2.271488141	5.32225E-14	1.46159E-11
Case t2 X Case t3	ENSECAG00000003774	IGHG1	2.157271251	6.11847E-14	1.64117E-11
Case t2 X Case t3	ENSECAG00000012179	RGS2	1.630506187	1.11621E-13	2.92599E-11
Case t2 X Case t3	ENSECAG00000010277	PPP1R15A	1.503818816	1.71201E-13	4.38808E-11
Case t2 X Case t3	ENSECAG00000008555	LTB	2.003356512	2.7128E-13	6.80205E-11
Case t2 X Case t3	ENSECAG00000019922	ADAMDEC1	2.355693585	4.53342E-13	1.11252E-10
Case t2 X Case t3	ENSECAG00000013692	IGL	2.639383224	5.58415E-13	1.34182E-10
Case t2 X Case t3	ENSECAG00000022059	MYC	1.555354914	6.40528E-13	1.50772E-10
Case t2 X Case t3	ENSECAG00000012527	ENSECAG00000012527	-2.532313598	1.00159E-12	2.31047E-10
Case t2 X Case t3	ENSECAG00000026913	ENSECAG0000	-2.617271225	1.27844E-12	2.89127E-10

		0026913			
Case t2 X Case t3	ENSECAG00000023399	GADD45G	1.493241798	1.93254E-12	4.28653E-10
Case t2 X Case t3	ENSECAG00000010847	PGLYRP1	2.151552793	3.15514E-12	6.8663E-10
Case t2 X Case t3	ENSECAG00000000436	FCN1	1.798933702	5.87134E-12	1.25407E-09
Case t2 X Case t3	ENSECAG00000015010	CYP4F3	1.828240309	6.89398E-12	1.44573E-09
Case t2 X Case t3	ENSECAG00000000400	AMICA1	1.667306579	7.61419E-12	1.56825E-09
Case t2 X Case t3	ENSECAG00000016610	CTGF	1.434782053	1.38568E-11	2.80394E-09
Case t2 X Case t3	ENSECAG00000016661	NR4A1	1.344950583	2.18034E-11	4.33586E-09
Case t2 X Case t3	ENSECAG00000023992	S100P	1.597944532	3.43324E-11	6.71169E-09
Case t2 X Case t3	ENSECAG00000018649	UCP2	1.503545561	4.06927E-11	7.82249E-09
Case t2 X Case t3	ENSECAG00000006561	TRBC2	1.474580243	4.90167E-11	9.26817E-09
Case t2 X Case t3	ENSECAG00000006674	RRAD	1.346281032	5.29021E-11	9.8415E-09
Case t2 X Case t3	ENSECAG00000000548	IGHM	1.636314299	5.48074E-11	1.00341E-08
Case t2 X Case t3	ENSECAG00000015657	TNNT1	1.4281332	1.20509E-10	2.17179E-08
Case t2 X Case t3	ENSECAG00000004763	XIRP1	1.293548826	1.94338E-10	3.44845E-08
Case t2 X Case t3	ENSECAG00000009742	S100A12	2.019360365	2.1197E-10	3.70434E-08
Case t2 X Case t3	ENSECAG00000010089	DUSP6	1.288609136	3.23354E-10	5.56652E-08
Case t2 X Case t3	ENSECAG00000007960	LIMD2	1.60180998	4.69878E-10	7.96997E-08
Case t2 X Case t3	ENSECAG00000019932	LGALS9C	1.413355658	5.05261E-10	8.44592E-08
Case t2 X Case t3	ENSECAG00000000419	TRAC	1.489322423	5.28441E-10	8.63161E-08
Case t2 X Case t3	ENSECAG00000020452	TNNC1	1.369704764	5.31337E-10	8.63161E-08
Case t2 X Case t3	ENSECAG00000010020	HBB	2.119562201	6.78278E-10	1.08656E-07
Case t2 X Case t3	ENSECAG00000014875	PTPRCAP	1.4345552	7.53092E-10	1.18988E-07
Case t2 X Case t3	ENSECAG00000016339	ADAMTS1	1.250351605	8.17222E-10	1.27376E-07
Case t2 X Case t3	ENSECAG00000024788	CORO1A	1.349284846	1.16008E-09	1.78404E-07
Case t2 X Case t3	ENSECAG00000007867	MYL2	1.316652752	1.4061E-09	2.13394E-07
Case t2 X Case t3	ENSECAG00000017410	MYL3	1.309558654	2.60811E-09	3.90675E-07
Case t2 X Case t3	ENSECAG00000018395	TYROBP	1.309718806	5.23159E-09	7.73605E-07
Case t2 X Case t3	ENSECAG00000020463	HSPB1	1.187957489	5.31064E-09	7.75354E-07
Case t2 X Case t3	ENSECAG00000015834	DMD	-1.201494777	5.69262E-09	8.20733E-07
Case t2 X Case t3	ENSECAG00000014552	ENSECAG00000014552	-1.223198618	6.2907E-09	8.95765E-07
Case t2 X Case t3	ENSECAG00000017799	MYO1G	1.308878054	6.98058E-09	9.81879E-07
Case t2 X Case t3	ENSECAG00000009192	SRGN	1.203957895	7.67809E-09	1.06698E-06
Case t2 X Case t3	ENSECAG00000009474	ENSECAG00000009474	1.97331704	8.21324E-09	1.12776E-06
Case t2 X Case t3	ENSECAG00000000763	CSRNP	1.311913009	8.31742E-09	1.12862E-06

		1			
Case t2 X Case t3	ENSECAG00000024888	CCL5	1.28672592	9.00109E-09	1.20719E-06
Case t2 X Case t3	ENSECAG00000016776	TNNI1	1.257187544	9.85638E-09	1.30671E-06
Case t2 X Case t3	ENSECAG00000009363	IRG1	1.854095132	1.0617E-08	1.39155E-06
Case t2 X Case t3	ENSECAG00000011254	SMG1	-1.260155461	1.43074E-08	1.85417E-06
Case t2 X Case t3	ENSECAG00000012148	NMES1	1.690723335	1.50469E-08	1.92834E-06
Case t2 X Case t3	ENSECAG00000002668	RHOB	1.148074246	1.77873E-08	2.2545E-06
Case t2 X Case t3	ENSECAG00000012572	ENSECAG00000012572	-1.183876651	1.93002E-08	2.41966E-06
Case t2 X Case t3	ENSECAG00000020638	NEB	-1.122713663	2.39429E-08	2.96943E-06
Case t2 X Case t3	ENSECAG00000023322	ID3	1.158760487	2.48422E-08	3.04819E-06
Case t2 X Case t3	ENSECAG00000019124	LILRA5	1.516827666	2.63168E-08	3.19514E-06
Case t2 X Case t3	ENSECAG00000001159	FOXS1	1.260488919	2.75603E-08	3.31126E-06
Case t2 X Case t3	ENSECAG00000001784	CD14	1.168974284	3.75389E-08	4.46364E-06
Case t2 X Case t3	ENSECAG00000016508	HK3	1.29768793	3.88245E-08	4.56941E-06
Case t2 X Case t3	ENSECAG00000018028	TLR2	1.357514574	4.04357E-08	4.71096E-06
Case t2 X Case t3	ENSECAG00000010918	SELP	1.305331929	5.07243E-08	5.85055E-06

Table 6. **DAVID clusters based on biological function for all of the comparisons.** The top 10 clusters are represented

Case t1 X Case t2

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
GOTERM_BP_FAT	GO:0001775~cell activation	39	5.28202E-16	371	4.954985584	9.76996E-13
GOTERM_BP_FAT	GO:0045321~leukocyte activation	34	1.96577E-14	371	5.12298679	3.46723E-11
GOTERM_BP_FAT	GO:0006955~immune response	57	1.69367E-13	371	3.012211415	2.98972E-10
SP_PIR_KEYWORDS	signal	139	2.08901E-12	463	1.776816747	2.95892E-09
GOTERM_BP_FAT	GO:0002684~positive regulation of immune system process	31	2.53556E-12	371	4.749462049	4.47435E-09
UP_SEQ_FEATURE	signal peptide	139	2.85648E-12	462	1.769368631	4.73535E-09
GOTERM_BP_FAT	GO:0046649~lymphocyte activation	28	5.73119E-12	371	5.130558453	1.01136E-08
GOTERM_BP_FAT	GO:0050865~regulation of cell activation	25	7.09486E-11	371	5.209087409	1.252E-07
SP_PIR_KEYWORDS	disulfide bond	124	1.03979E-10	463	1.761795582	1.47279E-07
GOTERM_BP_FAT	GO:0006952~defense response	48	1.47226E-10	371	2.845940438	2.59805E-07

Case t2 X Case t3

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
GOTERM_BP_FAT	GO:0010033~response to organic substance	15	2.20271E-05	73	3.855375905	0.03462474
GOTERM_MF_FAT	GO:0005539~glycosaminoglycan binding	7	5.79067E-05	64	10.14296875	0.07149988
GOTERM_BP_FAT	GO:0009617~response to bacterium	8	7.26185E-05	73	7.681453616	0.11410755
GOTERM_MF_FAT	GO:0030247~polysaccharide binding	7	9.82979E-05	64	9.220880682	0.12134468
GOTERM_MF_FAT	GO:0001871~pattern binding	7	9.82979E-05	64	9.220880682	0.12134468
GOTERM_BP_FAT	GO:0006357~regulation of transcription from RNA polymerase II promoter	14	0.00010476	73	3.568653313	0.16458541
INTERPRO	IPR004827:Basic-leucine zipper (bZIP) transcription factor	5	0.00012671	82	19.16589968	0.16021563
GOTERM_MF_FAT	GO:0008201~heparin binding	6	0.00013908	64	11.81705097	0.17165338
GOTERM_BP_FAT	GO:0032570~response to progesterone stimulus	4	0.00015443	73	37.0630137	0.24251377

GOTERM_BP_FAT	GO:0009991~response to extracellular stimulus	8	0.00016431	73	6.738729763	0.25802142
KEEG_PATHWAY	hsa04010:MAPK signaling pathway	9	2.29E-04	34	5.041308658	0.22493160

Table 7. Summary of pathways used by GSEA and their keywords

KEYWORD: glycolysis						
COMPARISON **T1 to T2**						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	35	-0.69490904	-1.9657172	0	0.001615509	**ENRICHED T2**
REACTOME_GLYCOLYSIS	21	-0.6476651	-1.9034623	0.002020202	0.004038773	**ENRICHED T2**
KEYWORD: glycolysis						
COMPARISON **T2 to T3**						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	35	0.49987125	1.6620822	0.02008032	0.045878693	**ENRICHED T2**
REACTOME_GLYCOLYSIS	21	0.4757577	1.5183991	0.033663366	0.07309489	**ENRICHED T2**
KEYWORD: aerobic respiration						
COMPARISON **T1 to T2**						
PATHWAYS	SIZE					
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	101					NO ENRICHMENT
MITOCHONDRIAL_RESPIRATORY_CHAIN	21					NO ENRICHMENT
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	54					NO ENRICHMENT
CELLULAR_RESPIRATION	18					NO ENRICHMENT
KEYWORD: aerobic respiration						
COMPARISON **T2 to T3**						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	

REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	101	-0.5509119	-1.4273653	0.23837209	0.8328028	NO ENRICHMENT
MITOCHONDRIAL_RESPIRATORY_CHAIN	21	-0.59782785	-1.3658537	0.22762646	0.48945075	NO ENRICHMENT
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	54	-0.6259762	-1.2918968	0.27237353	0.37738857	NO ENRICHMENT
CELLULAR_RESPIRATION	18	-0.5079385	-1.262809	0.26029962	0.3018512	**Suggestive ENRICHED T3**
KEYWORD: mitochondria biogenesis						
COMPARISON **T1 to T2**						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
MITOCHONDRION_ORGANIZATION_AND_BIOGENESIS	39	-0.45377496	-1.4908727	0.08102767	0.2633632	**Suggestive ENRICHED T2**
KEYWORD: mitochondria biogenesis						
COMPARISON **T2 to T3**						
PATHWAYS	SIZE					
MITOCHONDRION_ORGANIZATION_AND_BIOGENESIS	39					NO ENRICHMENT
KEYWORD: fatty acid oxidation						
COMPARISON **T1 to T2**						
PATHWAYS	SIZE					
FATTY_ACID_OXIDATION	15					NO ENRICHMENT
REACTOME_ACTIVATED_AMPK_STIMULATES_FATTY_ACID_OXIDATION_IN_MUSCLE	16					NO ENRICHMENT
KEYWORD: fatty acid oxidation						
COMPARISON **T2 to T3**						
PATHWAYS	SIZE					

FATTY_ACID_OXIDATION	15					NO ENRICHMENT
REACTOME_ACTIVATED_AMPK_STIMULATES_FATTY_ACID_OXIDATION_IN_MUSCLE	16					NO ENRICHMENT
KEYWORD: striated muscle contraction						
COMPARISON **T1 to T2**						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
REACTOME_STRIATED_MUSCLE_CONTRACTION	27	-0.43546376	-1.3614424	0.14078675	0.14367816	**ENRICHED T2**
STRIATED_MUSCLE_DEVELOPMENT	31	-0.432036	-1.4706602	0.08523908	0.09508197	NO ENRICHMENT
KEYWORD: striated muscle contraction						
COMPARISON **T2 to T3**						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
REACTOME_STRIATED_MUSCLE_CONTRACTION	27	-0.57616717	-1.7092016	0.038617887	0.047540985	**ENRICHED T3**
STRIATED_MUSCLE_DEVELOPMENT	31	-0.432036	-1.4706602	0.08523908	0.09508197	**ENRICHED T3**

Table 8. **Summary of pathways used by GSEA and their keywords.** ES: enrichment score. Core enrichment YES: Genes that contributed for the pathway to be significant

Glycolysis			
KEGG_GLYCOLYSIS_GLUCONEOGENESIS COMPARISON T2 to T3			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PKLR	pyruvate kinase, liver and RBC	0.13907583	Yes
HK3	hexokinase 3 (white cell)	0.25844055	Yes
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	0.339179	Yes
PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	0.38016167	Yes
ALDOC	aldolase C, fructose-bisphosphate	0.4313096	Yes
PGK1	phosphoglycerate kinase 1	0.48479673	Yes
PGAM4	phosphoglycerate mutase family member 4	0.49987125	Yes
PGM2	phosphoglucomutase 2	0.46578398	No
AKR1A1	aldo-keto reductase family 1, member A1 (aldehyde reductase)	0.45068592	No
ALDH3A2	aldehyde dehydrogenase 3 family, member A2	0.2655334	No
PFKP	phosphofructokinase, platelet	0.23819382	No
ENO1	enolase 1, (alpha)	0.19206505	No
DLD	dihydrolipoamide dehydrogenase	0.18905637	No
LDHA	lactate dehydrogenase A	0.18235655	No
TPI1	triosephosphate isomerase 1	0.1792393	No
ACSS2	acyl-CoA synthetase short-chain family member 2	0.16712666	No
ACSS1	acyl-CoA synthetase short-chain family member 1	0.15924266	No

PGM1	phosphoglucomutase 1	0.13391243	No
ALDH9A1	aldehyde dehydrogenase 9 family, member A1	0.10494192	No
DLAT	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	0.10598044	No
GPI	glucose phosphate isomerase	0.114747964	No
BPGM	2,3-bisphosphoglycerate mutase	0.09878541	No
GALM	galactose mutarotase (aldose 1-epimerase)	0.06806623	No
ALDH7A1	aldehyde dehydrogenase 7 family, member A1	0.07998766	No
HK1	hexokinase 1	0.0900982	No
PDHB	pyruvate dehydrogenase (lipoamide) beta	0.058411337	No
PFKM	phosphofructokinase, muscle	0.077250674	No
ALDOA	aldolase A, fructose-bisphosphate	0.09232678	No
ENO3	enolase 3 (beta, muscle)	0.10141759	No
PDHA1	pyruvate dehydrogenase (lipoamide) alpha 1	0.05898392	No
PGAM2	phosphoglycerate mutase 2 (muscle)	0.08422355	No
FBP2	fructose-1,6-bisphosphatase 2	0.09376936	No
PFKL	phosphofructokinase, liver	0.07848208	No
LDHB	lactate dehydrogenase B	0.0472478	No
ALDH1A3	aldehyde dehydrogenase 1 family, member A3	0.09059825	No

Glycolysis			
REACTOME_GLYCOLYSIS COMPARISON T2 to T3			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PKLR	pyruvate kinase, liver and RBC	0.2103427	Yes
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	0.33489734	Yes
ALDOC	aldolase C, fructose-bisphosphate	0.39475855	Yes
PGK1	phosphoglycerate kinase 1	0.4757577	Yes
PFKP	phosphofructokinase, platelet	0.14583541	No
ENO1	enolase 1, (alpha)	0.10057233	No
PPP2R5D	protein phosphatase 2, regulatory subunit B (B56), delta isoform	0.09365212	No
TPI1	triosephosphate isomerase 1	0.08818508	No
PFKFB4	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	0.059443966	No
GPI	glucose phosphate isomerase	0.00244271	No
PPP2CA	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	0.005339099	No
PFKM	phosphofructokinase, muscle	-0.08463839	No
ALDOA	aldolase A, fructose-bisphosphate	-0.059717953	No
PFKFB1	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	-0.034038655	No
ENO3	enolase 3 (beta, muscle)	-0.010541584	No
PPP2R1A	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), alpha isoform	-0.007139669	No
PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), beta isoform	0.013511842	No
PGAM2	phosphoglycerate mutase 2 (muscle)	0.029780176	No
PFKL	phosphofructokinase, liver	0.012962782	No
PFKFB2	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	0.040970795	No

PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	0.028644124	No

Glycolysis			
KEGG_GLYCOLYSIS_GLUCONEOGENESIS COMPARISON T1 to T2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
ALDH7A1	aldehyde dehydrogenase 7 family, member A1	-0.26898795	No
PFKL	phosphofructokinase, liver	-0.30118245	No
ALDH3A2	aldehyde dehydrogenase 3 family, member A2	-0.34663817	No
ALDH1A3	aldehyde dehydrogenase 1 family, member A3	-0.37494412	No
DLD	dihydrolipoamide dehydrogenase	-0.40559292	No
DLAT	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	-0.4231037	No
PKLR	pyruvate kinase, liver and RBC	-0.46189758	No
ALDH9A1	aldehyde dehydrogenase 9 family, member A1	-0.5486637	No
ACSS2	acyl-CoA synthetase short-chain family member 2	-0.60186356	No
HK1	hexokinase 1	-0.6437047	No
PFKM	phosphofructokinase, muscle	-0.62816906	No
PDHA1	pyruvate dehydrogenase (lipoamide) alpha 1	-0.66450095	No
FBP2	fructose-1,6-bisphosphatase 2	-0.6490245	No
PDHB	pyruvate dehydrogenase (lipoamide) beta	-0.6346623	No
ALDOA	aldolase A, fructose-bisphosphate	-0.62239015	No
BPGM	2,3-bisphosphoglycerate mutase	-0.61045015	No
LDHB	lactate dehydrogenase B	-0.66791993	Yes
PGM2	phosphoglucomutase 2	-0.64124805	Yes
PGM1	phosphoglucomutase 1	-0.6219788	Yes
AKR1A1	aldo-keto reductase family 1, member A1 (aldehyde reductase)	-0.5988273	Yes

ACSS1	acyl-CoA synthetase short-chain family member 1	-0.57401395	Yes
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	-0.55068946	Yes
ENO1	enolase 1, (alpha)	-0.537385	Yes
HK3	hexokinase 3 (white cell)	-0.51060873	Yes
GALM	galactose mutarotase (aldose 1-epimerase)	-0.48674762	Yes
PFKP	phosphofructokinase, platelet	-0.48884705	Yes
PGAM4	phosphoglycerate mutase family member 4	-0.45501506	Yes
TPI1	triosephosphate isomerase 1	-0.40542296	Yes
PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	-0.3517684	Yes
PGK1	phosphoglycerate kinase 1	-0.29785973	Yes
ENO3	enolase 3 (beta, muscle)	-0.24386156	Yes
PGAM2	phosphoglycerate mutase 2 (muscle)	-0.18923523	Yes
ALDOC	aldolase C, fructose-bisphosphate	-0.13250725	Yes
GPI	glucose phosphate isomerase	-0.0704595	Yes
LDHA	lactate dehydrogenase A	0.001490959	Yes

Glycolysis			
REACTOME_GLYCOLYSIS COMPARISON T1 to T2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	0.11386655	No
PFKFB1	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	0.045146883	No
PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), beta isoform	-0.12074023	No
PFKL	phosphofructokinase, liver	-0.15947343	No
PFKFB2	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	-0.16874024	No

PKLR	pyruvate kinase, liver and RBC	-0.32761362	No
PPP2R5D	protein phosphatase 2, regulatory subunit B (B56), delta isoform	-0.4235623	No
PPP2R1A	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), alpha isoform	-0.49327064	No
PFKM	phosphofructokinase, muscle	-0.48771518	No
ALDOA	aldolase A, fructose-bisphosphate	-0.5300232	No
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	-0.60436356	Yes
PPP2CA	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	-0.5705522	Yes
ENO1	enolase 1, (alpha)	-0.53083056	Yes
PFKP	phosphofructokinase, platelet	-0.52939093	Yes
PFKFB4	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	-0.46900293	Yes
TPI1	triosephosphate isomerase 1	-0.4099281	Yes
PGK1	phosphoglycerate kinase 1	-0.33125743	Yes
ENO3	enolase 3 (beta, muscle)	-0.2522874	Yes
PGAM2	phosphoglycerate mutase 2 (muscle)	-0.17203477	Yes
ALDOC	aldolase C, fructose-bisphosphate	-0.08814424	Yes
GPI	glucose phosphate isomerase	0.003153408	Yes

Aerobic respiration			
CELLULAR_RESPIRATION T2 to T3			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PPARGC1A	peroxisome proliferative activated receptor, gamma, coactivator 1, alpha	-0.047863778	No
NNT	nicotinamide nucleotide transhydrogenase	-0.34880745	No
NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	-0.47965235	Yes

SURF1	surfeit 1	-0.476851	Yes
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	-0.4457616	Yes
UQCRH	ubiquinol-cytochrome c reductase hinge protein	-0.4151305	Yes
SDHC	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa	-0.43732613	Yes
PDHB	pyruvate dehydrogenase (lipoamide) beta	-0.40865698	Yes
UQCRB	ubiquinol-cytochrome c reductase binding protein	-0.36510032	Yes
UQCRC2	ubiquinol-cytochrome c reductase core protein II	-0.32538426	Yes
CYCS	cytochrome c, somatic	-0.29229105	Yes
SLC25A14	solute carrier family 25 (mitochondrial carrier, brain), member 14	-0.28898168	Yes
SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	-0.23563294	Yes
ME3	malic enzyme 3, NADP(+)-dependent, mitochondrial	-0.19035383	Yes
ACO2	aconitase 2, mitochondrial	-0.13451342	Yes
SDHB	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	-0.05881529	Yes
UQCRC1	ubiquinol-cytochrome c reductase core protein I	0.015974713	Yes
OXA1L	oxidase (cytochrome c) assembly 1-like	0.08993779	Yes

Mitochondria biogenesis			
MITOCHONDRION ORGANIZATION AND BIOGENESIS T1 to T2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
BCL2L1	BCL2-like 1	-0.03421537	No
POLRMT	polymerase (RNA) mitochondrial (DNA directed)	-0.019665433	No
COX18	COX18 cytochrome c oxidase assembly homolog (S. cerevisiae)	-0.10877428	No
BCL2	B-cell CLL/lymphoma 2	-0.12678462	No

MSTO1	misato homolog 1 (Drosophila)	-0.2200554	No
MFN2	mitofusin 2	-0.2226482	No
FIS1	fission 1 (mitochondrial outer membrane) homolog (S. cerevisiae)	-0.22191599	No
OPA1	optic atrophy 1 (autosomal dominant)	-0.24119109	No
BID	BH3 interacting domain death agonist	-0.2729822	No
STARD3	START domain containing 3	-0.28719562	No
SLC25A15	solute carrier family 25 (mitochondrial carrier; ornithine transporter) member 15	-0.2996277	No
BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3	-0.31462362	No
TIMM50	translocase of inner mitochondrial membrane 50 homolog (S. cerevisiae)	-0.3731208	No
TFB2M	transcription factor B2, mitochondrial	-0.3718092	No
SLC25A4	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4	-0.39457473	No
TRNT1	tRNA nucleotidyl transferase, CCA-adding, 1	-0.39393988	No
MIPEP	mitochondrial intermediate peptidase	-0.4365217	Yes
NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa (NADH-coenzyme Q reductase)	-0.42774674	Yes
TIMM17B	translocase of inner mitochondrial membrane 17 homolog B (yeast)	-0.41404352	Yes
HSP90AA1	heat shock protein 90kDa alpha (cytosolic), class A member 1	-0.4044121	Yes
TIMM17A	translocase of inner mitochondrial membrane 17 homolog A (yeast)	-0.40928012	Yes
TIMM44	translocase of inner mitochondrial membrane 44 homolog (yeast)	-0.40298837	Yes
MPV17	MpV17 mitochondrial inner membrane protein	-0.391342	Yes
NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	-0.37916794	Yes
TFAM	transcription factor A, mitochondrial	-0.39670292	Yes

TOMM22	translocase of outer mitochondrial membrane 22 homolog (yeast)	-0.3812983	Yes
MTX2	metaxin 2	-0.38879213	Yes
MRPL12	mitochondrial ribosomal protein L12	-0.39989534	Yes
SLC25A1	solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1	-0.37414286	Yes
IFI6	interferon, alpha-inducible protein 6	-0.3401471	Yes
TAZ	tafazzin (cardiomyopathy, dilated 3A (X-linked); endocardial fibroelastosis 2; Barth syndrome)	-0.30889738	Yes
DNM1L	dynamin 1-like	-0.27195662	Yes
BAK1	BCL2-antagonist/killer 1	-0.26316497	Yes
NDUFS5	NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa (NADH-coenzyme Q reductase)	-0.23313986	Yes
TSPO	translocator protein (18kDa)	-0.21255793	Yes
NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	-0.19862464	Yes
GPX1	glutathione peroxidase 1	-0.1414416	Yes
TOMM34	translocase of outer mitochondrial membrane 34	-0.07798036	Yes
BAX	BCL2-associated X protein	0.004474507	Yes

Muscle structure/ contraction			
STRIATED MUSCLE DEVELOPMENT COMPARISON T2 to T3			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
TGFB1	transforming growth factor, beta 1 (Camurati-Engelmann disease)	0.020540724	No
MKL2	MKL/myocardin-like 2	0.035044942	No
TTN	titin	0.03964372	No

MYH11	myosin, heavy chain 11, smooth muscle	-0.001456962	No
MBNL1	muscleblind-like (Drosophila)	-0.057826404	No
HDAC9	histone deacetylase 9	-0.107838936	No
SVIL	supervillin	-0.17907901	No
MYF5	myogenic factor 5	-0.25892287	No
NRD1	nardilysin (N-arginine dibasic convertase)	-0.31160536	No
NOTCH1	Notch homolog 1, translocation-associated (Drosophila)	-0.3057011	No
MYL6	myosin, light chain 6, alkali, smooth muscle and non-muscle	-0.31764394	No
BOC	Boc homolog (mouse)	-0.30960748	No
CHRNA1	cholinergic receptor, nicotinic, alpha 1 (muscle)	-0.30100578	No
MYOZ1	myozenin 1	-0.35903817	No
MYL6B	myosin, light chain 6B, alkali, smooth muscle and non-muscle	-0.40682152	Yes
IFRD1	interferon-related developmental regulator 1	-0.38805324	Yes
CACNA1H	calcium channel, voltage-dependent, alpha 1H subunit	-0.37127852	Yes
ACTA1	actin, alpha 1, skeletal muscle	-0.40591496	Yes
IGFBP3	insulin-like growth factor binding protein 3	-0.402336	Yes
CACNB2	calcium channel, voltage-dependent, beta 2 subunit	-0.39274564	Yes
HDAC5	histone deacetylase 5	-0.37332374	Yes
SIRT2	sirtuin (silent mating type information regulation 2 homolog) 2 (S. cerevisiae)	-0.37655866	Yes
HDAC4	histone deacetylase 4	-0.35423258	Yes
MAPK12	mitogen-activated protein kinase 12	-0.32556358	Yes
TAZ	tafazzin (cardiomyopathy, dilated 3A (X-linked); endocardial fibroelastosis 2; Barth syndrome)	-0.31428343	Yes
BMP4	bone morphogenetic protein 4	-0.27068335	Yes
MYF6	myogenic factor 6 (herculin)	-0.22545144	Yes

TBX3	T-box 3 (ulnar mammary syndrome)	-0.18836586	Yes
MYOD1	myogenic differentiation 1	-0.1326932	Yes
MYEF2	myelin expression factor 2	-0.07235352	Yes
CSRP3	cysteine and glycine-rich protein 3 (cardiac LIM protein)	0.00832884	Yes

Muscle structure/ contraction			
REACTOME STRIATED MUSCLE CONTRACTION COMPARISON T2 to T3			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
TNNT2	troponin T type 2 (cardiac)	0.06958364	No
DMD	dystrophin (muscular dystrophy, Duchenne and Becker types)	0.046786375	No
VIM	vimentin	0.067855164	No
NEB	nebulin	0.093322285	No
MYL4	myosin, light chain 4, alkali; atrial, embryonic	-0.09171164	No
MYL1	myosin, light chain 1, alkali; skeletal, fast	-0.14351046	No
TPM4	tropomyosin 4	-0.14367516	No
TPM1	tropomyosin 1 (alpha)	-0.15416498	No
TMOD1	tropomodulin 1	-0.18672857	No
TPM2	tropomyosin 2 (beta)	-0.34792015	No
ACTN2	actinin, alpha 2	-0.36552334	No
MYH8	myosin, heavy chain 8, skeletal muscle, perinatal	-0.41172448	No
TNNC2	troponin C type 2 (fast)	-0.4413656	No
DES	desmin	-0.42408025	No
MYBPC2	myosin binding protein C, fast type	-0.45877194	No
TNNI2	troponin I type 2 (skeletal, fast)	-0.44600224	No
TCAP	titin-cap (telethonin)	-0.47418872	No
MYBPC1	myosin binding protein C, slow type	-0.48536965	No

TPM3	tropomyosin 3	-0.5268663	Yes
TNNC1	troponin C type 1 (slow)	-0.4928308	Yes
MYH3	myosin, heavy chain 3, skeletal muscle, embryonic	-0.42703176	Yes
MYL2	myosin, light chain 2, regulatory, cardiac, slow	-0.36065263	Yes
TNNT1	troponin T type 1 (skeletal, slow)	-0.29308853	Yes
TNNI3	troponin I type 3 (cardiac)	-0.22549802	Yes
MYL3	myosin, light chain 3, alkali; ventricular, skeletal, slow	-0.15566489	Yes
TNNI1	troponin I type 1 (skeletal, slow)	-0.078808434	Yes
MYH6	myosin, heavy chain 6, cardiac muscle, alpha (cardiomyopathy, hypertrophic 1)	0.008589001	Yes

Muscle structure/ contraction			
REACTOME COMPARISON T1 to T2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
MYL4	myosin, light chain 4, alkali; atrial, embryonic	0.038398545	No
TNNI3	troponin I type 3 (cardiac)	0.010624884	No
MYH3	myosin, heavy chain 3, skeletal muscle, embryonic	0.04085993	No
ACTN2	actinin, alpha 2	0.06784227	No
DES	desmin	0.079609826	No
MYBPC1	myosin binding protein C, slow type	0.07167236	No
DMD	dystrophin (muscular dystrophy, Duchenne and Becker types)	0.08182785	No
TNNT2	troponin T type 2 (cardiac)	0.032254517	No
MYH6	myosin, heavy chain 6, cardiac muscle, alpha (cardiomyopathy, hypertrophic 1)	0.035329115	No
NEB	nebulin	0.02225417	No
MYH8	myosin, heavy chain 8, skeletal muscle, perinatal	0.024042623	No
TCAP	titin-cap (telethonin)	-0.15751894	No

MYBPC2	myosin binding protein C, fast type	-0.26426032	No
TNNC1	troponin C type 1 (slow)	-0.27327535	No
TNNI1	troponin I type 1 (skeletal, slow)	-0.26841295	No
TNNC2	troponin C type 2 (fast)	-0.2917354	No
TMOD1	tropomodulin 1	-0.3492236	No
TNNT1	troponin T type 1 (skeletal, slow)	-0.39830953	Yes
TPM3	tropomyosin 3	-0.3812591	Yes
TNNI2	troponin I type 2 (skeletal, fast)	-0.37670282	Yes
MYL3	myosin, light chain 3, alkali; ventricular, skeletal, slow	-0.33295575	Yes
TPM4	tropomyosin 4	-0.31633145	Yes
MYL2	myosin, light chain 2, regulatory, cardiac, slow	-0.26498863	Yes
TPM1	tropomyosin 1 (alpha)	-0.24811675	Yes
VIM	Vimentin	-0.18421523	Yes
MYL1	myosin, light chain 1, alkali; skeletal, fast	-0.09504266	Yes
TPM2	tropomyosin 2 (beta)	0.005170855	Yes

Chapter 3

Gene Expression Profile in Type 1 Polysaccharide Storage Myopathy

Summary

Polysaccharide Storage Myopathy (PSSM) is a form of glycogen storage disease in horses, characterized by abnormal polysaccharide inclusions in skeletal muscle. PSSM1 is caused by a dominant gain of function mutation in the *GYS1* gene. While a gain of function in glycogen synthase explains the excessive muscle glycogen in PSSM1 horses (1.5 to 2 times the normal amount), the link between excessive muscle glycogen, abnormal polysaccharide and rhabdomyolysis during sub-maximal exercise is less clear. PSSM1 horses can metabolize glycogen and have a normal flux of metabolites through glycolysis during maximal exercise, yet these horses demonstrate exercise intolerance, painful muscle cramping, and rhabdomyolysis during sub-maximal exercise.

To evaluate the changes in muscle that lead to this energy deficit, muscle gene expression profiles before and after a controlled exercise trial were evaluated in PSSM1 cases and controls by RNASeq. Gene counts were determined by HTSeq, and EdgeR was used to calculate changes in gene expression between cases and controls at different time points (prior to the trial, and before and after exercise at the end of the trial). 201 genes were differentially expressed between cases and controls at time point one (pre-trial), 301 genes were differently expressed between cases and controls at time point 2 (pre-exercise, end of trial), and 803 genes were differentially expressed between cases and controls at time point 3 (post-exercise at the end of the trial). Pathway analysis performed by gene set enrichment analysis (GSEA) revealed enrichment in pathways involving mitochondria biogenesis, oxidative phosphorylation, fatty acid metabolism, glycogen and glucose metabolism. DAVID was used to cluster the top differentially expressed genes based on their functional annotation on each one of the comparisons performed. Clusters involved in inflammation were overrepresented.

Manual curation of pathways based on what is known about PSSM1 will enable a better annotation of the gene sets and is the next step to better understand the changes identified by the current analyses and the pathophysiology of this disease.

Introduction

Polysaccharide Storage Myopathy (PSSM) is a muscle glycogen storage disease in horses characterized by accumulation of excess and abnormal skeletal muscle glycogen and clinical myopathy[2]. A dominant mutation in the *GYS1* gene has been described as a causative mutation in one form of PSSM (PSSM1)[12,25]. This mutation occurs in the coding sequence of exon 6 and results in amino acid substitution from arginine to histidine leading to a gain-of-function in the skeletal muscle isoform of glycogen synthase[25]. It has been hypothesized that this amino acid substitution leads to poor regulation and increased activity of glycogen synthase, which is the rate limiting step on glycogen synthesis. The mutation may result in the enzyme being in a continuously active state, which in turn could be responsible for the accumulation of excessive glycogen and abnormal amylase-resistant, less branched polysaccharide in horses with type 1 PSSM[25].

PSSM horses have 1.5 to 2 times higher glycogen concentration in the muscle compared with controls, but have the ability to breakdown glycogen during exercise and produce lactate[21]. However, during submaximal exercise (which relies mostly on aerobic metabolism), PSSM horses show disruption of cellular energy metabolism and considerable metabolic stress, characterized by increased IMP levels in single muscle fibers and increased serum CK activity[23]. Muscle energy metabolism is closely regulated by the cytosolic ATP:ADP ratio. A decrease in this ratio stimulates oxidative metabolism, followed by myokinase reaction if ATP can't be restored. The myokinase reaction combines 2 ADP and produces ATP and adenine monophosphate (AMP), which will activate AMP deaminase and consequently increase levels of muscle inosine monophosphate (IMP). Increased levels of IMP in the muscle usually occur during intense exercise or under metabolic stress [23]. This IMP elevation after submaximal (~ 20 minute) exercise is comparable with horses performing at maximal exercise. Even with as little as 11 minutes of walk and trot, horses with PSSM have adenine nucleotide degradation in some fibers that resembles horses performing at twice the speed for almost 5 times as long. During light submaximal exercise, healthy horses should not have any

adenine depletion[23,21]. IMP accumulation in PSSM horses suggests that they have a skeletal muscle energy deficit with submaximal exercise. The mechanism underlying this energy deficit remains unknown.

Skeletal muscle of transgenic mice over-expressing *GYS1,* with resultant accumulation of excess glycogen, shows altered expression of > 700 genes; 98 which are directly involved in cellular energy metabolism, and 44 for which expression varies in either direct or inverse proportion to cellular glycogen content [74].

. From the genes whose expression varies with glycogen concentrations, 3 (besides *GYS1*) are involved in cellular metabolism, including hydroxysteroid (17-beta) dehydrogenase 7 (*HSD17b7*), UDP-glucose pyrophosphorylase 2 (*Ugp2*) and carbonic anhydrase 14 (Car14). Alteration in *Ugp2* expression is perhaps not surprising, as this gene encodes an enzyme important for mammalian glycogen metabolism. This enzyme transfers a glucose moiety from glucose-1-phosphate to MgUTP and forms UDP-glucose and MgPPi. The UDP-glucose is a direct precursor of glycogen in muscle and liver. Eleven of the genes with varied expression are involved in muscle and contractile proteins, including *Atp1β1* (ATPase, Na+/K+ Transport, Beta-1 Polypeptide), *ATP2α2* (ATPase, Ca++ transporting, cardiac muscle, slow twitch 2), myosin light and heavy chain family genes and others; and 14 other genes are involved in cell signaling [74].

An equine homemade oligonucleotide microarray has been created to evaluate gene expression profile in 7 PSSM1 and 6 control Norman Cob horses. PSSM horses had amylase-resistant polysaccharide in the muscle and were either homozygous or heterozygous for the *GYS1* mutation. The array included a total of 384 transcripts (50 probes of the mitochondrial genome and 334 probes of the nuclear genome) and 129 genes were found to be differently expressed between cases and controls. Thirteen genes involved in inflammatory pathways were up regulated in cases and six genes involved in mitochondrial function were down regulated. The top 10 up and down regulated genes (total of 20 genes) were also evaluated by quantitative RT-PCR. The expression levels of at least 4 of those genes were not in agreement between both methods. It is important to point out that this study was performed in a different breed of horse than the present

study and the number of oligo probes present in the array was very low and represents just a small portion of the transcriptome. Also a few of the microarray expression findings did not correlate with the quantitative RT-PCR expression results, raising questions about the quality of the probes[75].

While a gain of function in glycogen synthase (GS) likely explains the muscle biopsy phenotype in type 1 PSSM, the link between excessive muscle glycogen, abnormal polysaccharide and rhabdomyolysis during sub-maximal exercise is less clear. PSSM1 horses can metabolize glycogen and have a normal flux of metabolites through glycolysis during maximal exercise[21], yet these horses demonstrate exercise intolerance, painful muscle cramping, and rhabdomyolysis during sub-maximal (aerobic) exercise.

Carbohydrates and fatty acids are the two major substrates for aerobic metabolism. During the first 30 minutes of submaximal exercise most of the energy is provided by muscle glycogen[40]. Circulating FFA concentrations increase within 15 minutes of low-intensity exercise and uptake of short-chain and medium chain FFA across muscle membranes occurs down a concentration gradient by diffusion. Long-chain FFA are translocated across the muscle membrane by active transport (fatty acid transporters). Transport of FFA from the cytoplasm into mitochondria for β-oxidation is facilitated by an acylcarnitine transport system. A continuous supply of pyruvate from glucose or glycogen is essential for metabolism of FFA. Glycogen provides pyruvate which replenishes intermediates in the citric acid cycle (particularly oxaloacetate) allowing acetyl coenzyme A (CoA) derived from β-oxidation of FFA to enter the first step of the citric acid cycle[40].

The energy deficit observed during submaximal exercise in PSSM horses might be secondary to the constant activation of the GS enzyme or excessive glycogen stores, due to the gain of function mutation in *GYS1* gene. This scenario could be exacerbated by insulin release when consuming grain, which might be interpreted by the cell as an indication that glycogen breakdown and lipolysis are not necessary. Normally, the cell would sense an energy deficit and AMP kinase activation would stimulate oxidative

metabolism of fatty acid and carbohydrate by activation of pyruvate dehydrogenase. If pyruvate dehydrogenase is not fully activated adequate acetyl CoA for oxidative metabolism is not produced. If PSSM horses are on a grain diet, they have low plasma FFA concentrations, possibly due to suppression of lipolysis by high insulin. So fatty acid oxidation, also fails in supplying acetyl CoA. Also, PSSM horses on a grain diet have high muscle citrate concentrations making them potentially unable to generate sufficient acetyl CoA for muscle energy from carbohydrate or fat metabolism during submaximal exercise[30 108]. High citrate concentrations activate acetyl CoA carboxylase that converts acetyl CoA to malonyl CoA, which directs acetyl CoA away from the citric acid cycle and causes inhibition of carnitine palmytoyl transferase (CPT1), the key enzyme necessary to transport long-chain fatty acids into the mitochondria for β-oxidation[40]. Thus, the muscle cell's mechanism to senses an energy deficit might not be working properly due to constant up regulation of the glycogen synthase enzyme and/or excessive muscle glycogen and this process could be accentuated if horses are on a grain diet.

Current recommendations to control rhabdomyolysis in PSSM horses include regular exercise, increased turnout time, and a diet with low non-structural carbohydrates and high fat concentration. The provision of long-chain fatty acids in the diet of PSSM horses significantly decreases exercise intolerance and rhabdomyolysis with exercise[1 29]. Fat supplementation increases plasma FFA concentrations and the availability of fat for oxidation in skeletal muscle. It has been shown that the contribution of FFAs to energy production during low-intensity exercise is largely dependent on supply. Daily exercise may lower plasma insulin and increase plasma FFA concentrations, and training over time may enhance uptake of fatty acids into skeletal muscle and improve muscle oxidative capacity[40].

Whole genome next generation RNA sequencing is a high throughput sequencing technology of cDNA. It has several advantages over technologies for transcriptome analysis, such as microarray. It achieves base-pair level resolution and has a higher dynamic range of expression levels. Also, sequencing does not rely on probe design and previous genome knowledge and therefore is not limited to detecting known transcripts

and isoforms and permits de novo transcriptome assembly when a high quality reference genome or no reference genome is available. A few concerns have to be taken into consideration when performing RNA-seq, including how uniformly sequences from an entire transcript will be represented, how much sequence is required to detect and measure the transcripts with low abundance and appropriate downstream analysis since expression levels rely on proper mapping of sequencing reads to the corresponding reference genome or on their efficiency of de novo assembly [45-47,76]. Pathway analysis is then performed to gain insights into the underlying biology of differentially expressed genes at the function level[49].

The purpose of this study was to investigate whether dyregulation of gene expression in key metabolic pathways could be responsible for the energy deficit and rhabdomyolysis that occurs with exercise in type 1 PSSM horses. The expression of genes and gene pathways that are potentially altered in PSSM1 horses during subclinical rhabdomyolysis, and the altered expression patterns after training, were determined by quantifying and comparing gene expression using next generation sequencing of the skeletal muscle transcriptome (RNA-seq) in PSSM1 cases and controls.

Material and Methods

Study Design

Horses. Sample sizes needed to detect gene expression differences in cases and controls were estimated by $n > (\sigma 2 \log f)/ (z2\ 1-\alpha)$, ($\sigma 2$, variance in gene expression was estimated from data in other species[64], fold change = 2; and $\alpha=0.05$). Based on this calculation, a minimum of 5 biologic replicates in each group is necessary to detect differences in gene expression, consistent with standard recommendations[65]. Eight PSSM1 and 6 healthy control QHs; 9 mares and 5 geldings ranging in age from 1.5 to 14 years (mean 6.3 +/- 3.9 years) were utilized for the study. Prior to training, horses were unfit, and had a minimum of 90 days rest in a small paddock with no forced exercise. Results of muscle biochemistry for PSSM1 horses have been reported previously[30].

Collection of biopsy samples and training protocol. Prior to the trial, a high-starch diet that predisposes PSSM1 horses to moderate rhabdomyolysis[29] was introduced and this diet was consumed for the duration of the study. Baseline percutaneous gluteal muscle biopsy samples were collected for all horses prior to the start of treadmill exercise (time point **t1**)[20]. Biopsy samples consisted of at least 0.25 g of tissue from either the left or right gluteal muscle; samples were immediately frozen in liquid nitrogen and stored at -80˚C pending further analysis. Horses had up to a two-week period of acclimatization to treadmill training (**Figure 7**). After the acclimatization horses completed a 3-week training period. Daily exercise for control horses consisted of 4 minutes of walk, and then alternating intervals of 2 min walk (1.9 m/s) and 2 min trot (3.0 – 3.8 m/s) for 20 min.

Because individual PSSM1 horses have variable exercise tolerance, individual "exercise targets" of up to 20 min per day were established for each horse during the 2-week acclimatization period. The goal of exercise targets for PSSM1 horses was to perform enough exercise to detect subclinical rhabdomyolysis (based on serum CK) without causing overt signs of muscle stiffness[32]. Similar to control horses, exercise consisted of 4 min walk (1.9 m/s) + repeated 2 min intervals of walk + trot (~3 m/s). At the end of the training period (day 21), muscle biopsy samples were collected before (time point **t2**) and after submaximal exercise (time point **t3**). To evaluate for subclinical rhabdomyolysis, CK activity was assessed 4 hours after every exercise bout in PSSM1 cases and twice weekly in controls.

RNA extraction and sequencing

Total RNA was isolated from muscle tissue using Trizol reagent, quantified by Nanodrop spectrophotometry, and examined for integrity of RNA bands on an Agilent Bionalyzer. Library construction was performed using the Illumina HiSeq genome analyzer at the University of Minnesota Genomics Center (UMGC), using the TruSeq RNA V2 LT Illumina kit. First, the procedure purifies mRNA from total RNA, using polyA selection. The mRNA was then chemically fragmented and converted into single-stranded cDNA using random hexamer priming. Next, the second strand was generated to create double-stranded cDNA that was ready for the TruSeq library construction

workflow, where adapters were ligated to end of the fragments. Each library was sequenced (50 bp single end reads) at a rate of 40 million reads per sample. Individual libraries were sequenced in more than one lane so possible batch effects could be later identified[66].

Quality Control and Mapping

Fastq files created from sequencing were converted to fastq Sanger format. Prior to mapping, quality control of the reads was performed using FastQC program. QC reports were generated for every library and analyzed for per base sequence and per sequence quality score, per base and per sequence GC content, per base N content, sequence length distribution, duplication levels and overrepresented sequences and Kmer content. Reads with a quality score lower than 20 were trimmed. The reads were then mapped to the equine reference genome (EqCab2) using Bowtie2 and the splice-aware aligner Tophat2[67]. First, Tophat uses Bowtie to align reads to the reference and breaks sequences that Bowtie cannot align on its own into shorter segments. Bowtie does not allow large gaps between read alignments to the reference genome, so it does not align reads that contain introns. The short segments created by Tophat[67,77] are then aligned to the reference genome. When segments created from the same reads aligned to the genome far from each other, Tophat infers that the reads spans a splice junction and builds an index of splice sites in the transcriptome. A gene model annotation file (GTF; gene transfer format) from ensemble was provided for the analysis. A GTF file is a tab-delimited file describing genomic features and their location. The reads from the same individual (library) were mapped by lane and the generated BAM files were then merged for follow-up analysis. A BAM file is a binary version of a SAM file, which is a tab-delimited text file that contains sequence alignment data.

Non-normalized (raw) gene count

A non-normalized gene count (raw count) was performed, as this is the required input for differential expression analysis based on negative binomial distribution (edgeR). The BAM files originating from mapping with Tophat were sorted and indexed by Samtools. HTSeq-count (a python package) was used to count the reads that mapped to

each one of the equine genes present in the GTF file. Given a file with aligned sequencing reads and a list of genomic features, HTSeq counts how many reads map to each feature. In this case, the read counts are the total number of reads aligned to each gene. HTSeq has three overlap resolution modes that dictate how aligned reads that overlap more than one genomic feature are treated (union, intersection-strict, intersection-nonempty). A union-counting mode was used to count the reads. The output is a tab-delimited table of read counts for each gene listed in the genomic annotation set.

Differential expression analysis - Edge R

Gene differential expression analysis was performed on edgeR (Bioconductor R package[78]) based on negative binomial distribution since there is more variation in RNA-seq data that can be accounted for by the Poisson model. In RNA-seq data, the variance grows faster than the mean, which is known as overdispersion, so the mean and the variance are modeled separately. EdgeR works on a table of read counts where rows correspond to genes and columns to independent libraries containing the non-normalized count produced by HTSeq. This design matrix was created into an object (DGE list) in R. Adjustment for sequencing depth is always done automatically by EdgeR, since different libraries can have different sequencing depths and it can affect read counts.

Another important issue is dealing with the total RNA output in a flow cell. Genes that have a very high expression level can consume most of the library size and cause the remaining genes to be under sampled; this can make them look down-regulated unless this effect is controlled for in the analysis. This is potentially a concern with skeletal muscle where a few genes such as actin and myosin will likely encompass a large proportion of the RNA-seq reads. The calcNormFactor function normalizes for RNA composition using a trimmed mean of M-values (TMM) between each pair of samples. This creates the effective library size, which is the dataset used for downstream analysis. EdgeR considers that GC content and gene length have little effect on differential expression analysis since they do not change from sample to sample (the same genes will have the same length and GC content regardless of the sample). Then dispersion was estimated and a linear model fitted (GLM) to ask the questions of interest (**figure 8**).

Pathway Analysis

Pathway analysis can be performed from using both bottom-up and top-down approaches. Bottom up approaches focus on examining a preset list of genes (which have the largest difference between groups) to evaluate their biological function. Top-down approaches focus on a pathway or gene list that has been identified *a priori,* and typically test for association with all gene members in the pathway or list, with the association statistic summing over multiple genes to determine statistical significance. Both Bottom-up (DAVID) and top-down (GSEA) approaches were used in analysis of this data.

David

Following identification of genes differentially expressed, biologic themes, disease-gene relationships and metabolic pathways within the gene lists were identified with tools available from the DAVID bioinformatics resource (NIAID, http://david.abcc.ncifcrf.gov/). DAVID consists of an integrated biological knowledgebase and analytic tools aiming at systematically extracting biological meaning from large gene lists. This is an exploratory procedure more than a pure statistical solution. The goal of the analysis is to map a large number of interesting genes in a list to the associated biological annotation, and then statistically highlight the most over-represented (enriched) biological annotation out of thousands of linked terms and contents.

First a gene list is provided for the analysis (differently expressed genes). In the gene name batch viewer, a gene history report is provided and information about each gene can be accessed. The genes from the gene list are then cluster into groups according to their biological function. An enrichment score is provided and represents how important is this gene group in the gene list (the higher the level the greater the importance). Next, the functional annotation chart tool identifies the most relevant (over-expressed) biological terms with a given gene list. In this output, enriched functional annotations are divided in different categories which include term (e.g., transcription factor activity), number of genes in each category, percentage of genes involved in this

category from the total gene list, and a modified Fisher exact p-value. Fisher Exact is adopted to measure the gene-enrichment in annotation terms. It ranges from 0 to 1 (a p-value of zero represents perfect enrichment). Other tools are also available to visualize the same data in different formats (pathway map viewer, tabular format and others).

Gene Set Enrichment Analysis – GSEA

Traditional strategies for gene expression analysis have focused on identifying individual genes (single gene analysis) that exhibit differences between two states of interest. Although useful, they fail to detect biological processes, such as metabolic pathways and responses, that are distributed across an entire network of genes and subtle at the level of individual genes[72]. Cellular processes often affect sets of genes acting together. An increase of 20% in all genes encoding members of a metabolic pathway may dramatically alter the flux through the pathway and may be more important than a 20-fold increase in a single gene[72].

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (cases and controls for example). Gene sets are defined based on prior biological knowledge about the physiology of the disease of interest. The prior defined set of genes is chosen by the user, based on what is known about the pathological process of interest. It is a tool to interpret gene expression profile results, extracting meaning from their expression level and gaining insights into biological mechanisms[72]. Instead of a single gene analysis, GSEA calculates expression data at the level of gene sets. The goal of the program is to determine whether members of a gene set tend to occur at the top or bottom of the expression dataset, in which case the gene set is correlated with the phenotype of interest. Given an *a priori* defined set of genes *S* (e.g., genes encoding products in a metabolic pathway, located in the same cytogenetic band, or sharing the same GO category), the goal of GSEA is to determine whether the members of *S* are randomly distributed throughout *L* or primarily found at the top or bottom. We expect that sets related to the phenotypic distinction will tend to

show the latter distribution. GSEA applies Kolmogorov-Smirnof test to find asymmetrical distributions for defined blocks of genes in datasets whole distribution.

The program requires an expression dataset (created by the user). The first column contains information about genes (HUGO gene names in case of RNA-Seq data), the second column a description of the genes that is ignored by the program, followed by sample names (third column onwards). The rows are filled with expression value for each feature (gene) in each sample. Each column contains expression values for 1 sample. A phenotype label set (associates each sample to a phenotype) and gene sets which are exported from MSigDb or created by the user are also required. The expression data used in the present study was the read count number generated from HTSeq.

First GSEA calculates a rank for each one of the genes present in the expression dataset that are also present in the gene set list from a specific predefined pathway (list of genes expressed in a specific condition and their expression values). The genes are ranked based on their correlation with the phenotype and a ranked list is created (list L). Then an enrichment score (ES) is calculated and reflects the degree to which a set S (gene list from a specific pathway) is overrepresented at the extremes (top or bottom) of the entire ranked list L. The score is calculated by walking down the list L, increasing a running-sum statistic when a gene is encountered that is in the gene set *S* and decreasing it when a gene is encountered that is not in *S.* The magnitude of the increase or decrease in the running-sum statistic depends on the correlation of the gene with the phenotype. The enrichment score (ES) is the maximum deviation from zero encountered in the random walk; it corresponds to a weighted Kolmogorov–Smirnov-like statistic.

GSEA then estimates the significance level of the ES by using an empirical phenotype-based permutation test procedure that preserves the complex correlation structure of the gene expression data. 1000 phenotype label swapping permutations were performed to determine ES score significance. Estimated significance level to account for multiple hypothesis testing is also performed since a large number of gene sets are often evaluated. A normalized enrichment score (*NES*) is also performed to account for the size

of the set and allows for comparison of the results between different pathways. The proportion of false positives is calculated using false discovery rate (FDR) corresponding to each *NES*. The FDR is the estimated probability that a set with a given *NES* represents a false positive finding; it is computed by comparing the tails of the observed and null distributions for the *NES.* The leading edge subset of a gene set is the subset of members that contribute most to the ES.

The Molecular Signature Database (MSigDB) is a catalog of 1,325 gene sets, which are annotated gene sets for use with GSEA software. Gene sets for pathway analysis from the MSigDB database were selected by the use of keywords. First, general key words known to be important based on what is known about PSSM1 were used. Also, gene names or symbols can be used as keywords. The genes with the highest differential expression levels between cases and controls and genes which expression values were significantly different in mice with genetically modified muscle glycogen content due to disruption of the muscle *GYS1* genes in comparison to mice line GSL30 that hyper-accumulates glycogen in the muscle were also used as keywords.

The analysis described above was first performed using a ranked gene list created by GSEA using the genes that are on the expression dataset and also on the gene set file, and correlating them with the phenotype. Following this analysis, a customized ranked gene list was created using the list of the top differential genes for each time point. Those genes were ranked by EdgeR according to their log-fold change and FDR. The pathway analysis was then repeated, using the customized gene list.

GSEA calls all the pathways with FDR< 0.25 significant. Due to the low number of horses in this analysis instead of the typical significance FDR cut-off of < 0.25, FDRs < 0.35 were called suggestive of being significant.

Results

CK activity

Baseline CK activity was mildly elevated in PSSM1 cases (mean 617 U/L ± 82) and was significantly greater than baseline CK activity in controls (mean 255 U/L ± 34, p=0.007). Four hour post-exercise CK activity in PSSM horses was highest at the start of training and declined significantly over the 3 week training period. CK activity was significantly higher in PSSM cases than in controls throughout the training protocol (**Figure 9**).

RNA and sequencing quality controls

Total RNA integrity following isolation using an Agilent Technologies 2100 Bioanalyzer revealed a RNA Integrity Number (RIN) value greater than eight for all samples. An average Q score for base calling above the quality cutoff of 30 was observed for all lanes. The Q score is a PHRED score. A Q score of 30 equates to a 1/1000 probability that the base call is incorrect. All samples also met the sequencing criteria of 40 million reads/sample.

Mapping summary

Approximately 90-95% of the total reads per library mapped to the equine reference genome, and approximately 8% of the reads per library did not align uniquely to the reference genome. In this chapter the reads were mapped to the equine GTF file to generate non-normalized counts and this information was used for downstream analysis.

Differentially expressed gene analysis in PSSM1 cases and controls

Figure 8 shows the questions of interest that were asked on the gene expression analysis based on the study design. The differences representing exercise effect and training effect for controls were described in Chapter 2. Here the differences in gene expression between cases and controls at all three time points (pre-trial **t1**; pre-exercise **t2** and post-exercise **t3**), and cases at different time points, were evaluated. Also, a comparison between the effect of training and exercise in cases and controls was

performed (**double difference; Figure 8 F and G**). **Table 9** reports the number of genes that were differentially expressed with an FDR < 0.05 in each one of the comparisons. Those top genes were added to DAVID for functional annotation. **Table 10** to **table 16** shows the top 100 genes differently expressed between groups at different time points.

Pathway analysis

DAVID

The top genes from the differential expression analysis were imported to DAVID to find clusters of biological function without a priori knowledge. The top 10 most significant clusters for each one of the comparisons are summarized on **Table 17.** Clusters involved in inflammation were overrepresented. Genes differentially expressed between cases at time point 1 and cases at time point 2 were also involved in hypoxia, response to oxygen and myofibril assembly. Genes differentially expressed between cases at time point 2 and cases at time point 3 were also identified as being important for mitochondrion and actin cytoskeleton. Those clusters (mitochondrion and actin cytoskeleton) were not between the top 10 clusters but also had FDRs lower than 0.05 and are reported on **table 17**.

GSEA

Different from DAVID the pathways analyzed on GSEA were chosen prior to the analysis based on the criteria described in the methods. The keywords used were: glycogen, glycolysis, mitochondria metabolism, mitochondria biogenesis, inflammation, muscle contraction and structure, and the following gene names: *CYT-B* and *NADH* (both of which were top ranked on the differentially expressed analysis performed by EdgeR) and *UDPG2* and *Hsd17B7* (expression varies with glycogen levels in mice muscle). A significant number of pathways selected by the use of those keywords overlapped (e.g. CYT-B, NADH and mitochondria metabolism revealed similar pathways). **Table 18** summarizes all the pathways used in the analysis and the ones that were significantly different between time points. **Table 19** shows all the significant pathways and the genes

included in the gene set and the ones from the expression data that were the major contributors to the high and significant enrichment scores. A total of 10 keywords were used and 14 pathways were found to be enriched.

When evaluating fatty acid oxidation, enrichment was found at t2 in controls when compared to cases (**figure 10A**). No enrichment was found when comparing cases and controls at time points 1 and 3. Three different pathways (when using the keyword CYT-B) were enriched between cases and controls at t2. Those pathways are: reactome TCA cycle and respiratory electron transport (FDR 0.23), KEEG oxidative phosphorylation (FDR 0.21) and reactome respiratory electron transport (FDR 0.20). The use of NADH as a keyword revealed 2 pathways that were enriched in controls at t2 in comparison to cases (mitochondrial respiratory chain – suggestive with FDR 0.26 and mitochondrial membrane - FDR 0.22).

When evaluating glucose metabolism 2 pathways were enriched; both in cases when compared to controls at t1 using the key word glycolysis (KEEG glycolysis and gluconeogenesis FDR 0.009 and Reactome glycolysis FDR 0.09; **figure 10 B**). The use of the keyword glycogen generated 2 pathways from which glycogen metabolism was enriched in cases in comparison to controls at time point 1 (FDR 0.22) and enriched in cases at time point 1 in comparison to cases at time point 2 (FDR 0.11). The other generated pathway (glycogen breakdown) was not enriched in any of the time points.

Pathways using the word inflammation were not enriched in any of the comparisons performed. Four pathways involved in cellular components and organization were enriched in cases in comparison to controls at t1 (negative regulation of cellular component organization and biogenesis FDR 0.17; regulation of cytoskeleton organization and biogenesis FDR 0.21; actin cytoskeleton organization and biogenesis FDR 0.24; cytoskeleton organization and biogenesis FDR 0.22). Pathways generated with the key word muscle contraction were not enriched in any of the time points.

The findings described above were generated using a rank list created by GSEA using the genes present in the gene set and expression matrix and their correlation with the phenotype. The same analysis was performed using a pre-ranked file containing the top genes differentially expressed at each time point generated by EdgeR. Both analyses were consistent and generated the same enrichment results.

Discussion

Overall, alterations in the expression of genes and pathways involved in oxidative metabolism, glycolysis, mitochondrial biogenesis and fatty acid metabolism are of great importance in PSSM1.

The increase in CK activity in controls in comparison to cases demonstrated that subclinical rhabdomyolysis occurred at the beginning of the exercise trial. The CK activity in cases decreased significantly by the end of the exercise study, which is consistent with the improvement observed in PSSM1 horses with controlled regular exercise.

Hundreds of genes were found to be differentially expressed in cases and controls at different time points. Those transcripts were clustered based on their biological function using DAVID software. Several clusters were identified and the ones involved in inflammation were overrepresented. Clusters involved in hypoxia, response to oxygen, myofibril assembly, mitochondrion and actin cytoskeleton were also identified.

Several genes involved in mitochondrial oxidation and glucose metabolism were present in the list of top genes differentially expressed (in most cases those genes had a lower expression in cases in comparison to controls). While this method evaluates only the statistical significance of genes by itself, there are advantages to analyze sets of genes together according to their biological functions. GSEA considers all of the genes in an experiment. If the genes only change moderately it may be difficult to find significant

changes by looking at each gene separately. If, on the other hand, many genes belonging to the same gene set are changed, even small changes can become significant. Pathways involved in fatty acid oxidation, oxidative metabolism, mitochondrial biogenesis and glycolysis were found to be enriched by GSEA in several of the comparisons between cases and controls. Most of the available pathways are missing key elements (genes) that would allow us to understand the direction of those changes and how they affect the physiological process. For example, the fatty acid oxidation pathway was observed to be enriched in controls at time point 2 in comparison to cases. Enrichment in this case means that the program found a correlation between the gene counts from the expression dataset (experimental expression) that were also present in the gene set (pathway set), with the phenotype (in this case, time point2). It does not give an answer to which direction this correlation is going (if it is an up or down regulation in fat metabolism in controls in relation to cases at that time point). If key gene elements were not missing from this pathway, the direction of the change could be further evaluated by looking at the gene expression values of those genes in the dataset and their biological function.

This study has a few other limitations. It has been shown that gene expression in muscle can increase hours after exercise. Several myogenic and metabolic genes, for example, have a peak in gene expression 4 to 8 hours post-exercise[73]. The samples in this study were collected immediately after exercise, so it is possible that some genes would have a higher/lower level of expression if the biopsies were performed in a few hours after exercise. Although this might be a limitation of the study, muscle biopsy immediately after exercise enables the identification of differentially expressed early response genes that are rapidly induced by exercise. Also, the library prep kit used was specific for polyA selection (mRNA) so regulatory elements, such as miRNA, that might be important for changes in gene expression might have been missed.

The read counts performed in this study used the horse GTF file (from Ensemble) to map reads to genes, so this process relies on proper annotation of the reference genome. This is less of a problem in model species, like mice and human, but can be a challenge in species where the reference is not as well annotated like the horse. To

overcome this issue *de novo* assembly of the raw reads can be performed. This method does not rely on a reference genome and novel transcripts and isoforms can be identified. The *de novo* transcripts and isoforms can also be included in the differential expression analysis.

Also, hundreds of transcripts with significant expression levels were not annotated in the reference genome and had only ensemble IDs without associated gene names. Several of those IDs could not be converted using Biomart and DAVID software and had to be manually annotated by blasting their sequence to available databases. Also, to perform GSEA the equine transcript names have to be converted to HUGO gene names, which can be a challenge since a greater number of isoforms are present in the human database and the conversion is not always straightforward. A reference genome with better annotation and the use of a more consistent gene nomenclature between species would greatly improve this issue.

In addition, changes in gene expression not always indicate alterations in protein from a specific gene. Posttranslational modification of amino acids and phosphorylation of proteins might also interfere with protein behavior and activity level.

In conclusion, alterations in the expression of genes and pathways involved in oxidative metabolism, glycolysis, mitochondrial biogenesis and fatty acid metabolism are of great importance in PSSM1. The pathways generated by the analysis performed on this chapter with high enrichment score in different time points and conditions (case/control) were very consistent between analyses and are of great importance to determine which pathways should be manually curated for follow up analysis. Due to incomplete annotation of the equine reference genome and the publicly available gene sets for pathway analysis it is difficult to say with complete certainty how those pathways are influencing the energy deficit observed in those horses at a lower level (from the gene set to the list of genes in the expression dataset and their specific biological function in a pathway). Manual annotation/creation of pathways based on what is known about PSSM1

and the use of this new more specific pathway on GSEA is the next step to better understand the pathophysiology of this disease.

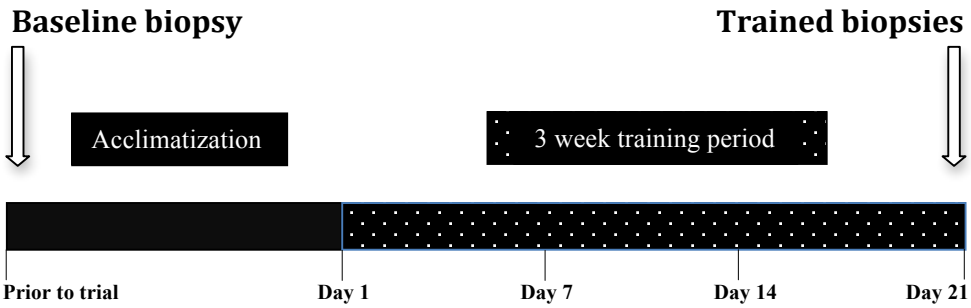
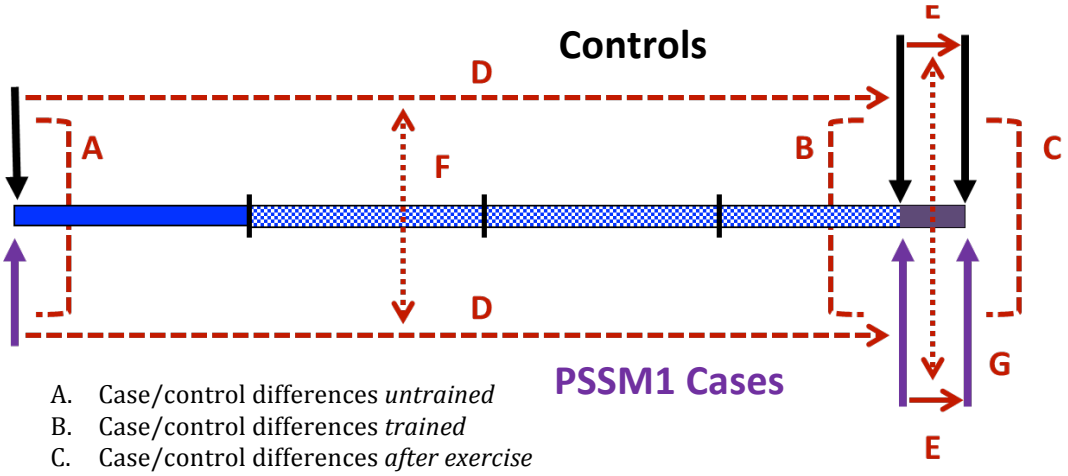


Figure 7. Sample collection protocol. Horses were acclimatized to the treadmill for 1 to 2 weeks. Horses trained 5 days/week for 3 weeks during the training period. On day 21 biopsies were taken before and after exercise



A. Case/control differences *untrained*
B. Case/control differences *trained*
C. Case/control differences *after exercise*
D. *Training affect* within group
E. *Exercise effect* within group
F. Differences in *training effect*
G. Differences in *exercise effect*

Figure 8. Comparisons performed by EdgeR between time points.

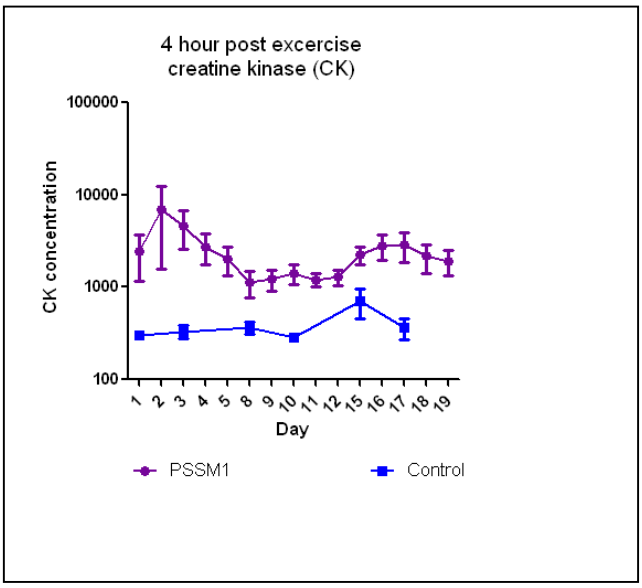


Figure 9. Post exercise CK activity during the 3 week training period. Serum CK activity (y-axis) was measured four hours post-exercise daily in cases and twice weekly in controls.

Figure 10A.

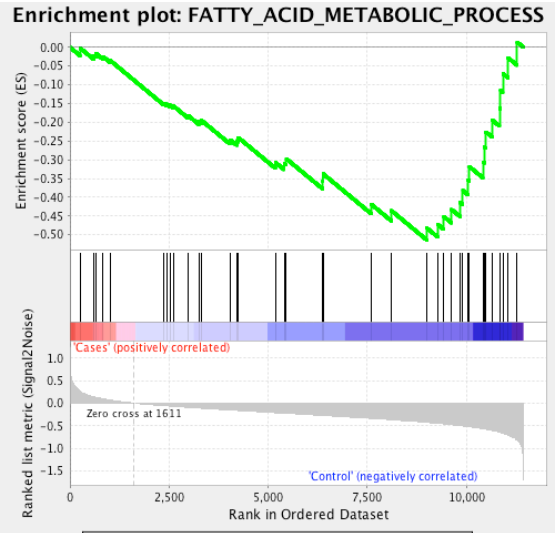


Figure 10B.

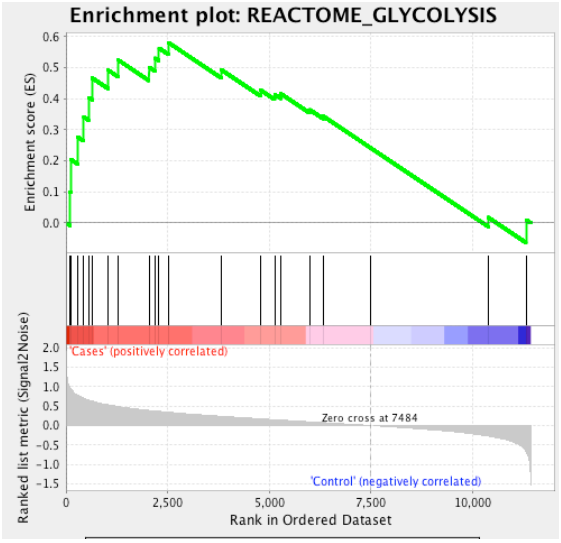


Figure 10.Case control analyses for 2 different pathways. Cases are represented on the left side of each plot and controls are represented on the right side of each plot. The top portion of the plots show the running ES for the gene set as the analysis walks down the ranked list. The score at the peak of the plot (the score furthest from 0.0, green line) is the ES for the gene set. Gene sets with a distinct peak at the beginning (picture on the right) or end of the ranked list (picture on the left) are generally the most interesting. Black bars illustrate the position of the gene sets belonging to each pathway. The ranked list metric shown in gray illustrates the correlation between the signal to noise values of all individually ranked genes according to the class labels (experimental conditions). A. Up-regulation of genes involved in fatty acid oxidation in controls. B. Up-regulation of genes involved in glycolysis in cases.

Table 9. Comparisons performed by EdgeR between time points

Comparison	Number of Genes Differentially Expressed
Case t1 X Case t2	124
Case t2 X Case t3	2781
Case t1 X Control t1	201
Case t2 X Control t2	303
Case t3 X Control t3	803
Case t1/Case t2 X Control t1/Control t2	236
Case t2/Case t3 X Control t2/Control t3	251

Table 10. **Comparisons performed between cases and controls at time point 1 (t1).** The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	**Ensembl Gene ID**	**Gene Name**	**logFC**	**PValue**	**FDR**
Case X Control t1	ENSECAG00000016536	MT-CYB	-6.44352971	4.18833E-50	4.83082E-46
Case X Control t1	ENSECAG00000000144	AKR1C1	7.147289214	1.55515E-44	8.96855E-41
Case X Control t1	ENSECAG00000021785	MHCB3	5.385977863	1.29323E-39	4.97205E-36
Case X Control t1	ENSECAG00000004995	TAL2	6.817512323	7.35227E-36	2.12003E-32
Case X Control t1	ENSECAG00000017642	GBP5	-4.02369594	2.37748E-22	5.48436E-19
Case X Control t1	ENSECAG00000019568	MYH13	3.891363384	2.50295E-21	4.81151E-18
Case X Control t1	ENSECAG00000009368	MICA	-3.359264206	2.47516E-18	4.07835E-15
Case X Control t1	ENSECAG00000000658	GPNMB	3.733808481	1.48044E-17	2.13442E-14
Case X Control t1	ENSECAG00000018002	FOSB	7.692903635	7.41621E-17	9.50428E-14
Case X Control t1	ENSECAG00000004135	AMPH	4.354576511	1.58206E-16	1.82475E-13
Case X Control t1	ENSECAG00000010841	AKR1C1	6.29745467	6.05027E-16	6.34399E-13
Case X Control t1	ENSECAG00000005658	MTND1	-3.249861369	1.16841E-14	1.12303E-11
Case X Control t1	ENSECAG00000000291	PLCG2	3.028463326	2.1164E-14	1.87773E-11
Case X Control t1	ENSECAG00000016867	CISH	-3.433100456	5.25637E-14	4.3305E-11
Case X Control t1	ENSECAG00000007258	IGHG1	4.050317454	3.29885E-13	2.5366E-10
Case X Control t1	ENSECAG00000021104	MYH6	-2.671436436	8.95666E-13	6.45663E-10
Case X Control t1	ENSECAG00000016737	SLC34A1	-3.620931828	3.51552E-12	2.27634E-09

Case X Control t1	ENSECAG00000008335	PDK4	-2.84733351	3.55246E-12	2.27634E-09
Case X Control t1	ENSECAG00000003845	HLA-A	2.612146261	8.80376E-12	5.34435E-09
Case X Control t1	ENSECAG00000003774	IGHG1	3.325283298	1.71127E-11	9.86887E-09
Case X Control t1	ENSECAG00000018476	TPX2	-2.621336512	1.92123E-11	1.05521E-08
Case X Control t1	ENSECAG00000007171	FGL1	2.910501933	2.70146E-11	1.4163E-08
Case X Control t1	ENSECAG00000014546	SLC7A4	2.446930383	6.58152E-11	3.30049E-08
Case X Control t1	ENSECAG00000003783	UCHL3	-2.745914544	2.64835E-10	1.27275E-07
Case X Control t1	ENSECAG00000008850	OSCP1	-2.499490042	3.59915E-10	1.6605E-07
Case X Control t1	ENSECAG00000017191	SPP1	2.326231695	4.59036E-10	2.03636E-07
Case X Control t1	ENSECAG00000003052	ZCCHC5	2.735353485	5.34485E-10	2.28324E-07
Case X Control t1	ENSECAG00000027336	RN5S1@	-2.615693126	5.87002E-10	2.41803E-07
Case X Control t1	ENSECAG00000009440	TYRP1	2.655802456	8.65097E-10	3.4407E-07
Case X Control t1	ENSECAG00000025060	MYH3	2.337624402	1.13628E-09	4.30731E-07
Case X Control t1	ENSECAG00000016634	RPS2	-2.72140274	1.15768E-09	4.30731E-07
Case X Control t1	ENSECAG00000024746	PDGFD	-2.552282187	2.42468E-09	8.73947E-07
Case X Control t1	ENSECAG00000015094	ATCAY	-3.197877066	2.72683E-09	9.53067E-07
Case X Control t1	ENSECAG00000022647	CHMP4C	-2.286552447	9.85543E-09	3.34331E-06
Case X Control t1	ENSECAG00000003944	GAPDH	-2.431083287	1.09558E-08	3.6104E-06
Case X Control t1	ENSECAG00000024656	KRT10	-3.107798868	1.48666E-08	4.7631E-06
Case X Control t1	ENSECAG00000013692	IGL	4.520148012	1.61978E-08	5.04934E-06
Case X Control t1	ENSECAG00000000591	TRAF3IP2	-2.271550418	1.91845E-08	5.72137E-06
Case X Control t1	ENSECAG00000000325	HLA-A	-2.062714602	1.93457E-08	5.72137E-06
Case X Control t1	ENSECAG00000012751	CCKBR	2.858969202	3.29454E-08	9.4998E-06
Case X Control t1	ENSECAG00000014515	CILP	2.093233726	3.8051E-08	1.06643E-05
Case X Control t1	ENSECAG00000010475	IL33	-2.244286718	3.8833E-08	1.06643E-05
Case X Control t1	ENSECAG00000027324	RNA28S1	-2.244126802	6.60847E-08	1.77261E-05
Case X Control t1	ENSECAG00000018724	ETNPPL	2.031390644	8.49557E-08	2.227E-05
Case X Control t1	ENSECAG00000013022	ACTC1	2.008164211	2.09216E-07	5.36245E-05
Case X Control t1	ENSECAG00000006674	RRAD	1.859979143	3.96495E-07	9.83687E-05
Case X Control t1	ENSECAG00000011854	CFB	2.180890095	4.00844E-07	9.83687E-05
Case X Control t1	ENSECAG00000009549	TNFRSF14	2.091431588	4.41017E-07	0.000105973
Case X Control t1	ENSECAG00000013384	STX3	2.232490446	4.75644E-07	0.000111961
Case X Control t1	ENSECAG00000018973	ENTHD1	1.958830583	6.56193E-07	0.000151371
Case X Control t1	ENSECAG00000014709	SDC1	2.08165221	7.12616E-07	0.000161163
Case X Control t1	ENSECAG00000008356	ZNF577	-1.974548323	9.38878E-07	0.00020825
Case X Control t1	ENSECAG00000021683	LGALS4	2.237321116	1.04161E-06	0.000226679
Case X Control t1	ENSECAG00000010717	MTMR7	-2.326509569	1.27425E-06	0.000272171
Case X Control t1	ENSECAG00000020599	HOXD10	-2.10608317	1.38602E-06	0.000290661
Case X Control t1	ENSECAG00000016560	CRABP2	1.959406783	1.55869E-06	0.000321035

Case X Control t1	ENSECAG00000019928	EQMHCC1	1.780506453	1.64908E-06	0.000333693
Case X Control t1	ENSECAG00000016531	KLHL6	1.972934904	1.95331E-06	0.000388439
Case X Control t1	ENSECAG00000024940	RNF207	1.831255331	2.17595E-06	0.000420123
Case X Control t1	ENSECAG00000013859	PIEZO2	2.057732939	2.18549E-06	0.000420123
Case X Control t1	ENSECAG00000013363	DACT2	2.417109589	2.69752E-06	0.000510053
Case X Control t1	ENSECAG00000022673	NSUN7	-2.260179178	2.81121E-06	0.000521
Case X Control t1	ENSECAG00000016873	CPEB1	2.004579871	2.84576E-06	0.000521
Case X Control t1	ENSECAG00000009625	IGHA1	2.915785796	3.41959E-06	0.000616275
Case X Control t1	ENSECAG00000016486	ARRDC2	-1.744349081	3.66517E-06	0.00065037
Case X Control t1	ENSECAG00000006669	ACSM3	-2.050104353	3.78966E-06	0.000657914
Case X Control t1	ENSECAG00000005924	TSKU	1.77721529	3.82177E-06	0.000657914
Case X Control t1	ENSECAG00000018254	UCHL1	1.725722072	4.02736E-06	0.000683111
Case X Control t1	ENSECAG00000019795	ARID5B	-1.732537184	4.27155E-06	0.00071403
Case X Control t1	ENSECAG00000001379	RBM17	-2.057950623	4.70524E-06	0.00077529
Case X Control t1	ENSECAG00000000943	PTGES	1.817464981	6.19082E-06	0.001005703
Case X Control t1	ENSECAG00000008177	FAM213B	1.683290456	6.37907E-06	0.001021892
Case X Control t1	ENSECAG00000006203	RAP1A	-2.175919083	7.12503E-06	0.001111212
Case X Control t1	ENSECAG00000015998	ABRA	-1.624812642	7.12933E-06	0.001111212
Case X Control t1	ENSECAG00000026307	U1	-2.216865645	7.69394E-06	0.001169701
Case X Control t1	ENSECAG00000023023	OSGIN1	2.220375433	7.70741E-06	0.001169701
Case X Control t1	ENSECAG00000013728	MFSD2A	-2.06339479	8.08346E-06	0.001210839
Case X Control t1	ENSECAG00000022841	NMRK2	1.728508498	9.17542E-06	0.001356786
Case X Control t1	ENSECAG00000025814	5S_rRNA	-2.227279299	9.88223E-06	0.001442805
Case X Control t1	ENSECAG00000026964	RPTN	-1.659554249	1.08637E-05	0.001565992
Case X Control t1	ENSECAG00000027377	5_8S_rRNA	-1.581700449	1.09975E-05	0.001565992
Case X Control t1	ENSECAG00000000847	SLC16A9	-1.958679506	1.11819E-05	0.001572837
Case X Control t1	ENSECAG00000005487	IGLC7	4.050972304	1.26291E-05	0.001754983
Case X Control t1	ENSECAG00000014817	RND2	1.713844404	1.35886E-05	0.00186584
Case X Control t1	ENSECAG00000022695	MAP1A	1.614876629	1.59054E-05	0.002158266
Case X Control t1	ENSECAG00000019318	HLA-B	1.715804399	1.71506E-05	0.00230018
Case X Control t1	ENSECAG00000000430	ACE2	2.129071935	1.79829E-05	0.002373281
Case X Control t1	ENSECAG00000017093	TPC3	-2.431903261	1.81072E-05	0.002373281
Case X Control t1	ENSECAG00000020210	RIC3	1.898956938	1.85068E-05	0.002398401
Case X Control t1	ENSECAG00000020112	LZTS1	-1.994545074	1.99774E-05	0.002554497
Case X Control t1	ENSECAG00000000548	IGHM	1.939893308	2.01543E-05	0.002554497
Case X Control t1	ENSECAG00000024790	ENSECAG00000024790	1.683436373	2.47102E-05	0.003097903
Case X Control t1	ENSECAG00000014255	SAG	-1.882965907	2.91525E-05	0.003615537
Case X Control t1	ENSECAG00000003702	CLEC4G	1.567178373	3.02035E-05	0.003706035
Case X Control t1	ENSECAG00000017655	PNP	1.871015908	3.20445E-05	0.00389054

Case X Control t1	ENSECAG00000013935	MTUS2	1.876429009	3.65482E-05	0.004391115
Case X Control t1	ENSECAG00000021702	PANX2	1.813050554	3.70445E-05	0.004404854
Case X Control t1	ENSECAG00000015212	MT3	-1.568927675	3.79474E-05	0.004466181
Case X Control t1	ENSECAG00000024596	CCP110	-1.735065924	4.13104E-05	0.004812866
Case X Control t1	ENSECAG00000010104	KCNIP2	1.574258528	4.884E-05	0.005583402

Table 11. **Comparisons performed between cases and controls at time point 2 (t2).** The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl Gene ID	Gene Name	logFC	PValue	FDR
Case X Control t2	ENSECAG00000016536	MT-CYB	-6.279302249	2.24967E-49	2.59477E-45
Case X Control t2	ENSECAG00000000144	AKR1C1	7.498775828	7.78211E-49	4.48794E-45
Case X Control t2	ENSECAG00000021785	MHCB3	5.466465229	5.14082E-40	1.97647E-36
Case X Control t2	ENSECAG00000005186	HMGB1	-7.594922843	2.18073E-30	6.28813E-27
Case X Control t2	ENSECAG00000017642	GBP5	-4.442999814	1.4022E-26	3.2346E-23
Case X Control t2	ENSECAG00000004995	TAL2	5.546112346	1.42193E-24	2.73342E-21
Case X Control t2	ENSECAG00000003944	GAPDH	-3.778484464	7.04629E-24	1.16103E-20
Case X Control t2	ENSECAG00000016634	RPS2	-3.928122718	2.89044E-23	4.16729E-20
Case X Control t2	ENSECAG00000019568	MYH13	3.66954906	3.74115E-22	4.79449E-19
Case X Control t2	ENSECAG00000005658	ENSECAG00000005658	-3.443461067	6.438E-20	7.42558E-17
Case X Control t2	ENSECAG00000004135	AMPH	4.676364894	2.33181E-18	2.44501E-15
Case X Control t2	ENSECAG00000009368	MICA	-3.327319611	2.67413E-17	2.57029E-14
Case X Control t2	ENSECAG00000017191	SPP1	2.822426359	1.80096E-16	1.59787E-13
Case X Control t2	ENSECAG00000003783	UCHL3	-3.347979538	4.62559E-16	3.81083E-13
Case X Control t2	ENSECAG00000007171	FGL1	3.650037825	6.2184E-16	4.78153E-13
Case X Control t2	ENSECAG00000010841	AKR1C1	6.330285964	6.72737E-16	4.8496E-13
Case X Control t2	ENSECAG00000014255	SAG	-3.583531518	2.77334E-15	1.88163E-12
Case X Control t2	ENSECAG00000001290	CABP1	-3.013107446	9.77689E-15	6.26481E-12
Case X Control t2	ENSECAG00000018002	FOSB	7.39043152	1.47993E-14	8.98394E-12
Case X Control t2	ENSECAG00000013022	ACTC1	3.094243431	5.51837E-14	3.18245E-11
Case X Control t2	ENSECAG00000006487	METTL21EP	-3.923824444	8.07671E-14	4.43604E-11
Case X Control t2	ENSECAG00000021104	MYH6	-2.771172537	1.49485E-13	7.83709E-11
Case X Control t2	ENSECAG00000023023	OSGIN1	3.689085107	1.70096E-13	8.52992E-11
Case X Control t2	ENSECAG00000013532	CCDC162P	4.416066954	2.92731E-13	1.40682E-10
Case X Control t2	ENSECAG00000002376	CHRNG	3.168669863	1.46386E-12	6.75368E-10
Case X Control t2	ENSECAG00000003845	HLA-A	2.70313616	7.84468E-12	3.48002E-09
Case X Control t2	ENSECAG00000001379	RBM17	-2.823289603	1.14444E-11	4.88888E-09
Case X Control t2	ENSECAG00000027324	RNA28S1	-2.552226197	1.44616E-11	5.95715E-09
Case X Control t2	ENSECAG00000011553	FAM65B	-2.704525614	2.52251E-11	1.00326E-08

Case X Control t2	ENSECAG00000005106	FIBIN	2.516312758	2.67293E-11	1.02765E-08
Case X Control t2	ENSECAG00000024790	ENSECAG00000024790	2.661669162	3.37989E-11	1.25754E-08
Case X Control t2	ENSECAG00000025060	MYH3	2.385463812	2.94947E-10	1.0631E-07
Case X Control t2	ENSECAG00000014709	SDC1	2.784183062	5.10633E-10	1.74944E-07
Case X Control t2	ENSECAG00000014546	SLC7A4	2.452785091	5.15702E-10	1.74944E-07
Case X Control t2	ENSECAG00000018254	UCHL1	2.297375204	1.11675E-09	3.68017E-07
Case X Control t2	ENSECAG00000006203	RAP1A	-2.632569233	1.60103E-09	5.12952E-07
Case X Control t2	ENSECAG00000000658	GPNMB	2.636052867	1.71319E-09	5.34052E-07
Case X Control t2	ENSECAG00000013859	PIEZO2	2.620545237	2.0226E-09	6.13912E-07
Case X Control t2	ENSECAG00000018973	ENTHD1	2.318907612	2.15024E-09	6.16078E-07
Case X Control t2	ENSECAG00000022647	CHMP4C	-2.355038577	2.17487E-09	6.16078E-07
Case X Control t2	ENSECAG00000009440	TYRP1	2.621522814	2.18998E-09	6.16078E-07
Case X Control t2	ENSECAG00000018724	ETNPPL	2.253282645	2.55345E-09	7.01226E-07
Case X Control t2	ENSECAG00000011389	YBX1	-2.181044386	3.01446E-09	8.08577E-07
Case X Control t2	ENSECAG00000016737	SLC34A1	-3.20096384	3.28038E-09	8.59908E-07
Case X Control t2	ENSECAG00000024746	PDGFD	-2.409219595	3.88485E-09	9.95731E-07
Case X Control t2	ENSECAG00000008957	NCF4	-2.249502939	7.4862E-09	1.87708E-06
Case X Control t2	ENSECAG00000023949	CCL2	2.765369967	1.0143E-08	2.48914E-06
Case X Control t2	ENSECAG00000024656	KRT10	-3.167159425	1.46659E-08	3.52409E-06
Case X Control t2	ENSECAG00000013836	PKLR	3.549210512	1.7143E-08	4.03526E-06
Case X Control t2	ENSECAG00000010847	PGLYRP1	-3.410625363	1.85431E-08	4.27752E-06
Case X Control t2	ENSECAG00000002023	SOD2	-2.181964564	1.90178E-08	4.30102E-06
Case X Control t2	ENSECAG00000007883	TLR7	-2.822542068	2.43933E-08	5.41063E-06
Case X Control t2	ENSECAG00000010475	IL33	-2.254000613	2.65612E-08	5.78032E-06
Case X Control t2	ENSECAG00000010928	MMP25	-2.677258333	3.86051E-08	8.13758E-06
Case X Control t2	ENSECAG00000014338	EGR1	2.251158964	3.88042E-08	8.13758E-06
Case X Control t2	ENSECAG00000011811	ENSECAG00000011811	3.456228166	4.75458E-08	9.70438E-06
Case X Control t2	ENSECAG00000021702	PANX2	2.459528637	4.79582E-08	9.70438E-06
Case X Control t2	ENSECAG00000024873	C1orf110	-2.675344743	5.22803E-08	1.03966E-05
Case X Control t2	ENSECAG00000013723	SLC7A8	-2.033176064	6.51522E-08	1.27367E-05
Case X Control t2	ENSECAG00000006610	CHMP4B	-2.117602302	8.64374E-08	1.63481E-05
Case X Control t2	ENSECAG00000000325	HLA-A	-2.011674521	8.64867E-08	1.63481E-05
Case X Control t2	ENSECAG00000010717	MTMR7	-2.560390598	8.78778E-08	1.63481E-05
Case X Control t2	ENSECAG00000017093	TPC3	-2.912117032	9.11345E-08	1.66849E-05
Case X Control t2	ENSECAG00000004837	CPN2	3.493356844	9.56025E-08	1.72294E-05
Case X Control t2	ENSECAG00000000430	ACE2	2.749075107	1.05576E-07	1.87341E-05
Case X Control t2	ENSECAG00000003052	ZCCHC5	2.158911512	1.7942E-07	3.10532E-05
Case X Control t2	ENSECAG00000014658	SCD	-2.02483265	1.80385E-07	3.10532E-05
Case X Control t2	ENSECAG00000019083	SORL1	-2.325454996	2.51682E-07	4.26897E-05
Case X Control t2	ENSECAG00000020485	PCOLCE2	2.085041003	2.59897E-07	4.3271E-05
Case X Control t2	ENSECAG00000013179	PCDH20	-2.482501425	2.64957E-07	4.3271E-05
Case X Control t2	ENSECAG00000023741	EGR3	2.356320618	2.66364E-07	4.3271E-05

Case X Control t2	ENSECAG00000016873	CPEB1	2.223759393	2.70948E-07	4.34044E-05
Case X Control t2	ENSECAG00000018071	PGPEP1L	2.044717576	2.85817E-07	4.51591E-05
Case X Control t2	ENSECAG00000013363	DACT2	2.751038848	3.05333E-07	4.75907E-05
Case X Control t2	ENSECAG00000000816	AMPD3	2.083843571	3.147E-07	4.83966E-05
Case X Control t2	ENSECAG00000018927	ANGPTL4	2.96149184	3.45228E-07	5.22737E-05
Case X Control t2	ENSECAG00000021683	LGALS4	2.309194745	3.48975E-07	5.22737E-05
Case X Control t2	ENSECAG00000013081	MMP9	-2.609361064	3.59156E-07	5.3109E-05
Case X Control t2	ENSECAG00000011292	MSR1	2.983476089	4.9353E-07	7.17552E-05
Case X Control t2	ENSECAG00000013129	TOP1MT	-2.384472561	4.97695E-07	7.17552E-05
Case X Control t2	ENSECAG00000013745	SNX10	-2.471013995	6.0592E-07	8.62801E-05
Case X Control t2	ENSECAG00000024845	PI16	1.985061006	7.13816E-07	0.000100404
Case X Control t2	ENSECAG00000024585	PDIA2	2.153246504	9.50219E-07	0.000130666
Case X Control t2	ENSECAG00000020122	SELL	-2.301853718	9.51616E-07	0.000130666
Case X Control t2	ENSECAG00000014566	WISP2	1.866462476	1.25947E-06	0.000170411
Case X Control t2	ENSECAG00000021044	C1QA	1.881313047	1.27062E-06	0.000170411
Case X Control t2	ENSECAG00000018101	FBLN1	1.897361598	1.28794E-06	0.000170748
Case X Control t2	ENSECAG00000008881	LYVE1	2.288043552	1.69352E-06	0.000221967
Case X Control t2	ENSECAG00000022673	NSUN7	-2.313687972	1.87672E-06	0.000243215
Case X Control t2	ENSECAG00000019557	LDLR	-1.833637839	1.95388E-06	0.000250401
Case X Control t2	ENSECAG00000024506	ARMC2	1.949170584	2.24416E-06	0.000284441
Case X Control t2	ENSECAG00000017295	PLSCR1	-2.010299438	2.33458E-06	0.000292685
Case X Control t2	ENSECAG00000002368	RPS15A	-1.646028284	2.68006E-06	0.000332385
Case X Control t2	ENSECAG00000003418	TMEM140	1.729741129	2.99131E-06	0.000364451
Case X Control t2	ENSECAG00000024939	ZNF385A	-1.855674162	3.0018E-06	0.000364451
Case X Control t2	ENSECAG00000012662	SEMA3A	-2.042304022	3.64226E-06	0.000437602
Case X Control t2	ENSECAG00000006877	CCDC88A	-1.696068494	3.93283E-06	0.000467642
Case X Control t2	ENSECAG00000010617	VGLL3	-1.716564128	4.3272E-06	0.000509285
Case X Control t2	ENSECAG00000020599	HOXD10	-1.974910747	5.6267E-06	0.000655538
Case X Control t2	ENSECAG00000019928	EQMHCC1	1.759599834	7.49464E-06	0.000864432

Table 12. **Comparisons performed between cases and controls at time point 3 (t3).** The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl Gene ID	Gene Name	logFC	PValue	FDR
Case X Control t3	ENSECAG00000016536	MT-CYB	-6.635778897	5.18186E-51	5.97675E-47
Case X Control t3	ENSECAG00000000144	AKR1C1	7.581446302	3.88594E-49	2.24102E-45
Case X Control t3	ENSECAG00000021785	MHCB3	5.62590826	3.6033E-41	1.38535E-37
Case X Control t3	ENSECAG00000018002	FOSB	9.001347733	2.06766E-28	5.96211E-25
Case X Control t3	ENSECAG00000017642	GBP5	-4.579445187	5.14037E-28	1.18578E-24
Case X Control t3	ENSECAG00000004995	TAL2	5.867184918	1.39859E-27	2.68857E-24
Case X Control t3	ENSECAG00000014338	EGR1	4.284435522	3.32853E-23	5.48447E-20
Case X Control t3	ENSECAG00000019568	MYH13	3.840934084	3.27601E-21	4.72319E-18

Case X Control t3	ENSECAG00000017191	SPP1	3.386270758	4.20501E-20	5.38895E-17
Case X Control t3	ENSECAG00000003816	JUNB	3.679581243	5.51781E-19	5.78568E-16
Case X Control t3	ENSECAG00000007171	FGL1	4.068153714	5.47622E-19	5.78568E-16
Case X Control t3	ENSECAG00000003845	HLA-A	3.366774392	6.62635E-18	6.36903E-15
Case X Control t3	ENSECAG00000014260	FOS	3.548851284	6.91942E-17	6.13912E-14
Case X Control t3	ENSECAG00000002376	CHRNG	3.633129519	8.78185E-17	6.75265E-14
Case X Control t3	ENSECAG00000010841	AKR1C1	6.47631027	8.50924E-17	6.75265E-14
Case X Control t3	ENSECAG00000013532	CCDC162P	5.104513221	4.97956E-16	3.58964E-13
Case X Control t3	ENSECAG00000004135	AMPH	4.229939468	3.14748E-15	2.13547E-12
Case X Control t3	ENSECAG00000009958	DUSP1	3.051552372	5.86287E-15	3.7568E-12
Case X Control t3	ENSECAG00000009012	IER2	2.99571758	1.04755E-13	6.35916E-11
Case X Control t3	ENSECAG00000009368	MICA	-2.86857496	1.23206E-13	7.10531E-11
Case X Control t3	ENSECAG00000001471	GPR34	3.828829237	1.81416E-13	9.96406E-11
Case X Control t3	ENSECAG00000005658	ENSECAG00000005658	-3.063160068	3.61924E-13	1.89747E-10
Case X Control t3	ENSECAG00000022125	TNNT2	2.879859339	4.5041E-13	2.25871E-10
Case X Control t3	ENSECAG00000016737	SLC34A1	-3.910548985	5.52101E-13	2.65331E-10
Case X Control t3	ENSECAG00000023023	OSGIN1	3.476517088	8.37482E-13	3.7152E-10
Case X Control t3	ENSECAG00000024746	PDGFD	-2.995524959	8.18832E-13	3.7152E-10
Case X Control t3	ENSECAG00000009549	TNFRSF14	2.967933031	1.57554E-12	6.73047E-10
Case X Control t3	ENSECAG00000014255	SAG	-3.038288155	6.27267E-12	2.58389E-09
Case X Control t3	ENSECAG00000021104	MYH6	-2.572751007	3.43649E-11	1.36678E-08
Case X Control t3	ENSECAG00000006487	METTL21EP	-3.503777449	4.44989E-11	1.71083E-08
Case X Control t3	ENSECAG00000010475	IL33	-2.742057167	5.28151E-11	1.96506E-08
Case X Control t3	ENSECAG00000007258	IGHG1	3.648616776	6.92867E-11	2.49735E-08
Case X Control t3	ENSECAG00000005106	FIBIN	2.443854996	8.67946E-11	3.0336E-08
Case X Control t3	ENSECAG00000018973	ENTHD1	2.492035286	9.65526E-11	3.27541E-08
Case X Control t3	ENSECAG00000024790	ENSECAG00000024790	2.574779674	1.87641E-10	6.18357E-08
Case X Control t3	ENSECAG00000023949	CCL2	2.873320551	2.3713E-10	7.59739E-08
Case X Control t3	ENSECAG00000014546	SLC7A4	2.357022832	6.11204E-10	1.9053E-07
Case X Control t3	ENSECAG00000027336	RN5S1@	-2.891549632	7.3992E-10	2.24585E-07
Case X Control t3	ENSECAG00000004180	HSPA6	3.148146505	1.04582E-09	3.09294E-07
Case X Control t3	ENSECAG00000010818	EGR2	3.186271221	1.0846E-09	3.12745E-07
Case X Control t3	ENSECAG00000017436	TREM1	3.030376365	1.6899E-09	4.75398E-07
Case X Control t3	ENSECAG00000018927	ANGPTL4	3.477820765	2.10855E-09	5.79049E-07
Case X Control t3	ENSECAG00000000816	AMPD3	2.363568246	2.93935E-09	7.88429E-07
Case X Control t3	ENSECAG00000014709	SDC1	2.641603183	3.22558E-09	8.45541E-07
Case X Control t3	ENSECAG00000024585	PDIA2	2.631610623	3.31911E-09	8.50724E-07
Case X Control t3	ENSECAG00000018254	UCHL1	2.225856638	3.6492E-09	9.14996E-07
Case X Control t3	ENSECAG00000003192	FTL	2.207436704	4.18826E-09	1.02782E-06
Case X Control t3	ENSECAG00000003783	UCHL3	-2.570787919	5.76725E-09	1.38582E-06
Case X Control t3	ENSECAG00000013723	SLC7A8	-2.181099464	7.44586E-09	1.75266E-06
Case X Control t3	ENSECAG00000017203	BTG2	2.25980575	8.24568E-09	1.90211E-06

Case X Control t3	ENSECAG00000001249	SOCS3	3.823063104	8.73047E-09	1.97445E-06
Case X Control t3	ENSECAG00000009625	IGHA1	3.165246837	1.13354E-08	2.51428E-06
Case X Control t3	ENSECAG00000003774	IGHG1	2.716488322	1.24802E-08	2.70675E-06
Case X Control t3	ENSECAG00000010717	MTMR7	-2.771382509	1.28897E-08	2.70675E-06
Case X Control t3	ENSECAG00000014566	WISP2	2.166916818	1.29072E-08	2.70675E-06
Case X Control t3	ENSECAG00000000291	PLCG2	2.18157265	1.37321E-08	2.82099E-06
Case X Control t3	ENSECAG00000005487	IGLC7	4.700151347	1.41857E-08	2.82099E-06
Case X Control t3	ENSECAG00000017093	TPC3	-3.197628571	1.39512E-08	2.82099E-06
Case X Control t3	ENSECAG00000023387	FRRS1	-2.091911067	1.46868E-08	2.87114E-06
Case X Control t3	ENSECAG00000020971	RND1	2.745166689	2.33916E-08	4.49665E-06
Case X Control t3	ENSECAG00000013179	PCDH20	-2.733958905	2.69842E-08	5.10223E-06
Case X Control t3	ENSECAG00000015261	IL18	2.450718838	3.22823E-08	6.00555E-06
Case X Control t3	ENSECAG00000022673	NSUN7	-2.706267053	3.89368E-08	7.12852E-06
Case X Control t3	ENSECAG00000012122	DLK1	-2.379683331	4.04331E-08	7.2868E-06
Case X Control t3	ENSECAG00000017807	NAV2	-2.103500163	4.13115E-08	7.33057E-06
Case X Control t3	ENSECAG00000011560	GSTO1	2.068696309	4.74128E-08	8.28575E-06
Case X Control t3	ENSECAG00000012662	SEMA3A	-2.446785456	4.86256E-08	8.37085E-06
Case X Control t3	ENSECAG00000014817	RND2	2.177342543	5.21172E-08	8.84E-06
Case X Control t3	ENSECAG00000027377	5_8S_rRNA	-2.305542088	5.65391E-08	9.45105E-06
Case X Control t3	ENSECAG00000025060	MYH3	2.107864354	7.71801E-08	1.27171E-05
Case X Control t3	ENSECAG00000019928	EQMHCC1	2.008779116	1.21112E-07	1.96748E-05
Case X Control t3	ENSECAG00000024656	KRT10	-2.963530191	1.42263E-07	2.27897E-05
Case X Control t3	ENSECAG00000022151	MATK	2.087146689	1.51962E-07	2.401E-05
Case X Control t3	ENSECAG00000013129	TOP1MT	-2.478771156	1.60062E-07	2.49481E-05
Case X Control t3	ENSECAG00000011292	MSR1	3.052642042	1.6581E-07	2.54994E-05
Case X Control t3	ENSECAG00000003944	GAPDH	-2.24167016	1.79046E-07	2.71726E-05
Case X Control t3	ENSECAG00000010617	VGLL3	-2.008365623	1.8969E-07	2.84141E-05
Case X Control t3	ENSECAG00000008177	FAM213B	1.997585248	2.03508E-07	3.0093E-05
Case X Control t3	ENSECAG00000013022	ACTC1	2.17580528	2.16101E-07	3.13427E-05
Case X Control t3	ENSECAG00000016873	CPEB1	2.251923627	2.17394E-07	3.13427E-05
Case X Control t3	ENSECAG00000011811	ENSECAG00000011811	3.466505017	2.42284E-07	3.44599E-05
Case X Control t3	ENSECAG00000018395	TYROBP	2.022794251	2.4499E-07	3.44599E-05
Case X Control t3	ENSECAG00000014500	ART5	1.942982689	2.66377E-07	3.70168E-05
Case X Control t3	ENSECAG00000001290	CABP1	-1.981930974	2.97134E-07	4.07205E-05
Case X Control t3	ENSECAG00000011068	CHI3L1	2.360708432	3.04924E-07	4.07205E-05
Case X Control t3	ENSECAG00000018071	PGPEP1L	2.077078008	3.04863E-07	4.07205E-05
Case X Control t3	ENSECAG00000023650	ABAT	-2.215541168	3.07151E-07	4.07205E-05
Case X Control t3	ENSECAG00000010846	DPP8	-1.885136449	3.36683E-07	4.39906E-05
Case X Control t3	ENSECAG00000024055	JUN	1.959029012	3.39446E-07	4.39906E-05
Case X Control t3	ENSECAG00000013425	IGFBP-5	-1.864301348	3.77614E-07	4.83933E-05
Case X Control t3	ENSECAG00000019318	HLA-B	2.0412217	4.07691E-07	5.11121E-05
Case X Control t3	ENSECAG00000022647	CHMP4C	-2.02518771	4.04176E-07	5.11121E-05

Case X Control t3	ENSECAG00000016942	NUDT14	2.051385558	4.18035E-07	5.18453E-05
Case X Control t3	ENSECAG00000025814	5S_rRNA	-2.714435664	4.4694E-07	5.48405E-05
Case X Control t3	ENSECAG00000000074	METRNL	2.05544234	5.46588E-07	6.37614E-05
Case X Control t3	ENSECAG00000000430	ACE2	2.568639899	5.45137E-07	6.37614E-05
Case X Control t3	ENSECAG00000003052	ZCCHC5	2.079159785	5.43414E-07	6.37614E-05
Case X Control t3	ENSECAG00000003418	TMEM140	1.849839114	5.47284E-07	6.37614E-05
Case X Control t3	ENSECAG00000004837	CPN2	3.313827519	5.36784E-07	6.37614E-05
Case X Control t3	ENSECAG00000010339	TLR4	1.914316263	5.93723E-07	6.848E-05

Table 13. **Comparisons performed between cases at time point 1 (t1) and cases at time point 2 (t2).** The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl Gene ID	Gene Name	logFC	PValue	FDR
Case t1 X Case t2	ENSECAG00000011484	EPS8L2	-4.78128379	1.23738E-52	1.4272E-48
Case t1 X Case t2	ENSECAG00000007809	OR2T27	-3.143150421	5.35865E-30	3.09033E-26
Case t1 X Case t2	ENSECAG00000016846	ENSECAG00000016846	-2.598850406	2.99273E-22	1.15061E-18
Case t1 X Case t2	ENSECAG00000020221	KNDC1	-3.2791048	6.42218E-22	1.85184E-18
Case t1 X Case t2	ENSECAG00000026913	ENSECAG00000026913	-2.10756729	1.96691E-18	4.53727E-15
Case t1 X Case t2	ENSECAG00000024589	ENSECAG00000024589	-2.311686808	3.89025E-17	7.47837E-14
Case t1 X Case t2	ENSECAG00000013836	PKLR	2.298607192	4.91875E-16	8.1047E-13
Case t1 X Case t2	ENSECAG00000019552	SLCO1C1	2.33545454	3.87642E-15	5.58883E-12
Case t1 X Case t2	ENSECAG00000013781	KIAA1211	-1.697206356	3.77175E-14	4.83371E-11
Case t1 X Case t2	ENSECAG00000006267	NR1D1	-1.346418978	7.26095E-11	7.61343E-08
Case t1 X Case t2	ENSECAG00000019557	LDLR	-1.388341109	6.80243E-11	7.61343E-08
Case t1 X Case t2	ENSECAG00000014500	ART5	1.271377618	2.75373E-10	2.64679E-07
Case t1 X Case t2	ENSECAG00000021096	NRXN2	-1.564598167	4.40512E-10	3.90836E-07
Case t1 X Case t2	ENSECAG00000024790	ENSECAG00000024790	1.322693497	2.19685E-09	1.80989E-06
Case t1 X Case t2	ENSECAG00000020485	PCOLCE2	1.296580385	2.54646E-09	1.95806E-06
Case t1 X Case t2	ENSECAG00000027324	RNA28S1	1.36718335	4.25282E-09	3.06575E-06
Case t1 X Case t2	ENSECAG00000023949	CCL2	1.515983122	1.90284E-08	1.29102E-05
Case t1 X Case t2	ENSECAG00000016634	RPS2	-1.170961299	9.40957E-08	6.02944E-05
Case t1 X Case t2	ENSECAG00000008710	FCER1A	1.588999592	1.08387E-07	6.57966E-05
Case t1 X Case t2	ENSECAG00000007161	CYP2F1	1.294710352	1.40488E-07	8.10194E-05
Case t1 X Case t2	ENSECAG00000002962	ADIPOQ	1.238907368	1.63498E-07	8.97993E-05
Case t1 X Case t2	ENSECAG00000014226	ENSECAG00000014226	1.062158811	2.28842E-07	0.000119976
Case t1 X Case t2	ENSECAG00000016225	TMEM144	1.665037785	2.63014E-07	0.000126209
Case t1 X Case t2	ENSECAG00000025328	RNA45S5	1.056872193	2.67001E-07	0.000126209
Case t1 X Case t2	ENSECAG00000027524	ENSECAG00000027524	1.058301003	2.73558E-07	0.000126209
Case t1 X Case t2	ENSECAG00000013022	ACTC1	1.097707766	3.73807E-07	0.000165827
Case t1 X Case t2	ENSECAG00000021019	CPXM1	1.102435473	4.71363E-07	0.000201359

Case t1 X Case t2	ENSECAG00000020878	MNDA	1.104024877	4.97943E-07	0.000205117
Case t1 X Case t2	ENSECAG00000013687	SMG1	-1.110581229	1.55494E-06	0.000618435
Case t1 X Case t2	ENSECAG00000027056	ENSECAG00000027056	0.982814572	1.89724E-06	0.000729427
Case t1 X Case t2	ENSECAG00000016560	CRABP2	1.00124932	2.22157E-06	0.000826569
Case t1 X Case t2	ENSECAG00000019220	PROX1	-1.1209383	3.81444E-06	0.001374867
Case t1 X Case t2	ENSECAG00000020059	FGF4	1.222608401	3.99579E-06	0.001396591
Case t1 X Case t2	ENSECAG00000016782	MYO1F	-0.960312958	4.84248E-06	0.001595806
Case t1 X Case t2	ENSECAG00000017431	MYOZ2	0.916496424	4.74411E-06	0.001595806
Case t1 X Case t2	ENSECAG00000027624	ENSECAG00000027624	0.956400782	5.02439E-06	0.001609758
Case t1 X Case t2	ENSECAG00000008916	ANKRD2	0.942918425	5.80386E-06	0.001805296
Case t1 X Case t2	ENSECAG00000021324	NR6A1	-1.022308695	5.94774E-06	0.001805296
Case t1 X Case t2	ENSECAG00000016582	SSPO	-1.118591246	7.19798E-06	0.002128757
Case t1 X Case t2	ENSECAG00000017807	NAV2	-0.916963886	8.5914E-06	0.002416907
Case t1 X Case t2	ENSECAG00000024660	DPF3	-0.943473031	8.5434E-06	0.002416907
Case t1 X Case t2	ENSECAG00000027699	ENSECAG00000027699	0.869780299	1.27308E-05	0.003496134
Case t1 X Case t2	ENSECAG00000026887	TSPAN8	0.871818503	1.39415E-05	0.003739571
Case t1 X Case t2	ENSECAG00000013517	LCOR	-0.902060461	1.5827E-05	0.004148821
Case t1 X Case t2	ENSECAG00000009126	NOS1	-0.849016463	2.32812E-05	0.005967233
Case t1 X Case t2	ENSECAG00000013728	MFSD2A	1.092417355	2.42829E-05	0.006088682
Case t1 X Case t2	ENSECAG00000015909	MYBPHL	0.965982026	2.52362E-05	0.006193059
Case t1 X Case t2	ENSECAG00000013913	PCK2	1.011936261	3.0714E-05	0.007115326
Case t1 X Case t2	ENSECAG00000018927	ANGPTL4	1.315852475	3.00746E-05	0.007115326
Case t1 X Case t2	ENSECAG00000018973	ENTHD1	0.882669224	3.0845E-05	0.007115326
Case t1 X Case t2	ENSECAG00000011254	SMG1	-0.881040155	3.25071E-05	0.007332647
Case t1 X Case t2	ENSECAG00000019769	PPARGC1B	-0.845040345	3.30586E-05	0.007332647
Case t1 X Case t2	ENSECAG00000008881	LYVE1	1.081690146	4.2435E-05	0.009063802
Case t1 X Case t2	ENSECAG00000023163	ENSECAG00000023163	-0.823719561	4.24087E-05	0.009063802
Case t1 X Case t2	ENSECAG00000009390	SEPT1	0.935311453	4.85474E-05	0.009823606
Case t1 X Case t2	ENSECAG00000023118	MCC	-0.839393141	4.82161E-05	0.009823606
Case t1 X Case t2	ENSECAG00000024143	MYO7A	-0.976939527	4.7703E-05	0.009823606
Case t1 X Case t2	ENSECAG00000001101	SESN1	0.791257883	5.67922E-05	0.01092791
Case t1 X Case t2	ENSECAG00000010266	NFAT5	-0.831615204	5.68471E-05	0.01092791
Case t1 X Case t2	ENSECAG00000019083	SORL1	-0.930650876	5.52259E-05	0.01092791
Case t1 X Case t2	ENSECAG00000000171	NPAS2	-0.988864073	6.25025E-05	0.011818092
Case t1 X Case t2	ENSECAG00000022544	KMT2C	-0.803892872	6.43492E-05	0.011971036
Case t1 X Case t2	ENSECAG00000019111	CD163	0.874990518	7.04406E-05	0.012499409
Case t1 X Case t2	ENSECAG00000023280	FADS1	-0.910205273	6.8483E-05	0.012499409
Case t1 X Case t2	ENSECAG00000024267	TFRC	-0.788703171	6.993E-05	0.012499409
Case t1 X Case t2	ENSECAG00000007663	GIMAP5	0.924855529	7.46506E-05	0.013045752
Case t1 X Case t2	ENSECAG00000018324	SLC43A2	0.794473572	7.76775E-05	0.013372127
Case t1 X Case t2	ENSECAG00000007047	ASPN	0.891294062	7.94172E-05	0.013470556
Case t1 X Case t2	ENSECAG00000007582	GIMAP7	0.915221414	8.58579E-05	0.014146927

Case t1 X Case t2	ENSECAG00000023076	GPCPD1	0.777738691	8.55911E-05	0.014146927
Case t1 X Case t2	ENSECAG00000006043	PHC3	-0.826927649	8.81954E-05	0.014327404
Case t1 X Case t2	ENSECAG00000000040	IKZF2	-0.842827447	9.06308E-05	0.014518552
Case t1 X Case t2	ENSECAG00000017294	SNTB1	-0.771538188	9.34903E-05	0.014771462
Case t1 X Case t2	ENSECAG00000009759	TNRC6B	-0.783049732	0.000100611	0.015681693
Case t1 X Case t2	ENSECAG00000017238	MTC1	0.893145088	0.000104608	0.016087346
Case t1 X Case t2	ENSECAG00000023387	FRRS1	-0.771461128	0.000108041	0.016396608
Case t1 X Case t2	ENSECAG00000020410	FREM2	-0.840703902	0.000113946	0.017068175
Case t1 X Case t2	ENSECAG00000015076	INO80D	-0.841658102	0.00012068	0.017845165
Case t1 X Case t2	ENSECAG00000008221	S100A4	0.777938675	0.000147727	0.021298474
Case t1 X Case t2	ENSECAG00000016173	TK1	0.898468848	0.000147266	0.021298474
Case t1 X Case t2	ENSECAG00000004773	TET2	-0.814472475	0.000154296	0.021971001
Case t1 X Case t2	ENSECAG00000024549	KIAA2018	-0.761033967	0.000161431	0.022706626
Case t1 X Case t2	ENSECAG00000017015	MYH11	-0.766792538	0.000165872	0.023050243
Case t1 X Case t2	ENSECAG00000005666	ZBED6	-0.755511021	0.000173707	0.023851587
Case t1 X Case t2	ENSECAG00000002206	SPRY3	-0.938833481	0.000191771	0.026022189
Case t1 X Case t2	ENSECAG00000012155	GAN	-0.814082401	0.000197192	0.026446642
Case t1 X Case t2	ENSECAG00000023650	ABAT	-0.872861552	0.000204686	0.027136222
Case t1 X Case t2	ENSECAG00000009535	FCER1G	0.801428029	0.000213226	0.027947086
Case t1 X Case t2	ENSECAG00000013723	SLC7A8	-0.744908856	0.000215895	0.027979051
Case t1 X Case t2	ENSECAG00000021790	LNPEP	-0.75209736	0.000222891	0.028564707
Case t1 X Case t2	ENSECAG00000017042	ENSECAG00000017042	0.854906711	0.000226468	0.028704213
Case t1 X Case t2	ENSECAG00000020053	SLC4A4	-0.792815614	0.000230056	0.028831319
Case t1 X Case t2	ENSECAG00000023632	RGS10	0.741984139	0.00023247	0.028831319
Case t1 X Case t2	ENSECAG00000018191	GUCY1A2	-0.818695963	0.000244669	0.030021411
Case t1 X Case t2	ENSECAG00000017946	KMT2D	-0.73014834	0.000258841	0.031425966
Case t1 X Case t2	ENSECAG00000000762	DDI2	-0.754614254	0.000272737	0.032230268
Case t1 X Case t2	ENSECAG00000016683	SMYD1	-0.718670775	0.000273848	0.032230268
Case t1 X Case t2	ENSECAG00000017427	OTOR	0.974489221	0.000270382	0.032230268
Case t1 X Case t2	ENSECAG00000000237	DGKH	-0.819461561	0.0002859	0.033308769
Case t1 X Case t2	ENSECAG00000007668	ENSECAG00000007668	0.925416797	0.000289727	0.033417098

Table 14. **Comparisons performed between cases at time point t2 (t2) and cases at time point 3 (t3).** The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl Gene ID	Gene Name	logFC	PValue	FDR
Case t2 X Case t3	ENSECAG00000014260	FOS	7.598763208	1.1264E-170	1.2992E-166
Case t2 X Case t3	ENSECAG00000014338	EGR1	7.033589317	7.7632E-161	4.477E-157
Case t2 X Case t3	ENSECAG00000018002	FOSB	8.638430603	2.0631E-146	7.932E-143
Case t2 X Case t3	ENSECAG00000003816	JUNB	4.473073737	4.89699E-81	1.41205E-77
Case t2 X Case t3	ENSECAG00000017203	BTG2	4.275525933	2.09277E-72	4.8276E-69
Case t2 X Case t3	ENSECAG00000014232	NR4A2	4.130895224	5.90984E-57	1.13607E-53
Case t2 X Case t3	ENSECAG00000010560	MGSA	4.774409938	1.12126E-56	1.84752E-53
Case t2 X Case t3	ENSECAG00000011486	ATF3	3.418073922	1.55216E-51	2.23783E-48
Case t2 X Case t3	ENSECAG00000010818	EGR2	3.999060604	4.51143E-50	5.78165E-47
Case t2 X Case t3	ENSECAG00000009012	IER2	3.304921184	6.50951E-49	7.50807E-46
Case t2 X Case t3	ENSECAG00000009958	DUSP1	3.09207054	4.50384E-46	4.72248E-43
Case t2 X Case t3	ENSECAG00000004180	HSPA6	3.755188486	4.94437E-45	4.75237E-42
Case t2 X Case t3	ENSECAG00000024993	CYR61	3.087759541	1.82075E-43	1.61542E-40
Case t2 X Case t3	ENSECAG00000010613	KLF4	2.830283968	1.69338E-37	1.39511E-34
Case t2 X Case t3	ENSECAG00000001249	SOCS3	3.838014794	1.60919E-34	1.23736E-31
Case t2 X Case t3	ENSECAG00000024055	JUN	2.576022754	3.60964E-33	2.6021E-30
Case t2 X Case t3	ENSECAG00000013836	PKLR	-5.039210367	6.11038E-33	4.14571E-30
Case t2 X Case t3	ENSECAG00000019552	SLCO1C1	-5.987065861	3.30454E-32	2.11748E-29
Case t2 X Case t3	ENSECAG00000020971	RND1	3.001285982	6.22398E-32	3.77829E-29
Case t2 X Case t3	ENSECAG00000016611	KLF2	2.367888106	9.14709E-28	5.27513E-25
Case t2 X Case t3	ENSECAG00000007242	DUSP2	4.011343384	3.77467E-27	2.07319E-24
Case t2 X Case t3	ENSECAG00000011488	MIP-2BETA	3.905463879	4.3806E-27	2.29663E-24
Case t2 X Case t3	ENSECAG00000020059	FGF4	-4.04051323	2.29056E-25	1.14867E-22
Case t2 X Case t3	ENSECAG00000005487	IGLC7	4.043096642	6.73026E-24	3.23445E-21
Case t2 X Case t3	ENSECAG00000017436	TREM1	2.788982413	3.46108E-22	1.5968E-19
Case t2 X Case t3	ENSECAG00000011496	NFKBIZ	2.123171438	5.58381E-22	2.47706E-19
Case t2 X Case t3	ENSECAG00000011895	ILT11B	2.629998058	3.49189E-20	1.49168E-17
Case t2 X Case t3	ENSECAG00000013457	CSF3R	2.168765643	1.38393E-18	5.70079E-16
Case t2 X Case t3	ENSECAG00000013081	MMP9	2.386807876	2.43589E-18	9.68813E-16
Case t2 X Case t3	ENSECAG00000004066	APOLD1	2.118307517	4.09667E-18	1.57503E-15
Case t2 X Case t3	ENSECAG00000013781	KIAA1211	-2.520995413	6.05905E-18	2.25436E-15
Case t2 X Case t3	ENSECAG00000013730	ID1	1.818135687	1.84767E-17	6.65969E-15
Case t2 X Case t3	ENSECAG00000023949	CCL2	2.009780927	2.03138E-17	7.09998E-15
Case t2 X Case t3	ENSECAG00000020122	SELL	2.095629557	1.59054E-16	5.39566E-14
Case t2 X Case t3	ENSECAG00000001471	GPR34	2.451200089	2.2118E-16	7.28884E-14

Case t2 X Case t3	ENSECAG00000004312	IL1RN	2.20639216	6.33413E-16	2.02938E-13
Case t2 X Case t3	ENSECAG00000025128	PIWIL2	-2.795650277	8.42879E-16	2.62751E-13
Case t2 X Case t3	ENSECAG00000023741	EGR3	1.829703695	3.47678E-15	1.0553E-12
Case t2 X Case t3	ENSECAG00000026998	LRRC10B	-2.702773129	8.86041E-15	2.62041E-12
Case t2 X Case t3	ENSECAG00000007258	IGHG1	2.507466372	1.56053E-14	4.49979E-12
Case t2 X Case t3	ENSECAG00000003072	C8orf4	1.640009076	3.07419E-14	8.64822E-12
Case t2 X Case t3	ENSECAG00000009625	IGHA1	2.271488141	5.32225E-14	1.46159E-11
Case t2 X Case t3	ENSECAG00000003774	IGHG1	2.157271251	6.11847E-14	1.64117E-11
Case t2 X Case t3	ENSECAG00000012179	RGS2	1.630506187	1.11621E-13	2.92599E-11
Case t2 X Case t3	ENSECAG00000010277	PPP1R15A	1.503818816	1.71201E-13	4.38808E-11
Case t2 X Case t3	ENSECAG00000008555	LTB	2.003356512	2.7128E-13	6.80205E-11
Case t2 X Case t3	ENSECAG00000019922	ADAMDEC1	2.355693585	4.53342E-13	1.11252E-10
Case t2 X Case t3	ENSECAG00000013692	IGL	2.639383224	5.58415E-13	1.34182E-10
Case t2 X Case t3	ENSECAG00000022059	MYC	1.555354914	6.40528E-13	1.50772E-10
Case t2 X Case t3	ENSECAG00000012527	ENSECAG00000012527	-2.532313598	1.00159E-12	2.31047E-10
Case t2 X Case t3	ENSECAG00000026913	ENSECAG00000026913	-2.617271225	1.27844E-12	2.89127E-10
Case t2 X Case t3	ENSECAG00000023399	GADD45G	1.493241798	1.93254E-12	4.28653E-10
Case t2 X Case t3	ENSECAG00000010847	PGLYRP1	2.151552793	3.15514E-12	6.8663E-10
Case t2 X Case t3	ENSECAG00000000436	FCN1	1.798933702	5.87134E-12	1.25407E-09
Case t2 X Case t3	ENSECAG00000015010	CYP4F3	1.828240309	6.89398E-12	1.44573E-09
Case t2 X Case t3	ENSECAG00000000400	AMICA1	1.667306579	7.61419E-12	1.56825E-09
Case t2 X Case t3	ENSECAG00000016610	CTGF	1.434782053	1.38568E-11	2.80394E-09
Case t2 X Case t3	ENSECAG00000016661	NR4A1	1.344950583	2.18034E-11	4.33586E-09
Case t2 X Case t3	ENSECAG00000023992	S100P	1.597944532	3.43324E-11	6.71169E-09
Case t2 X Case t3	ENSECAG00000018649	UCP2	1.503545561	4.06927E-11	7.82249E-09
Case t2 X Case t3	ENSECAG00000006561	TRBC2	1.474580243	4.90167E-11	9.26817E-09
Case t2 X Case t3	ENSECAG00000006674	RRAD	1.346281032	5.29021E-11	9.8415E-09
Case t2 X Case t3	ENSECAG00000000548	IGHM	1.636314299	5.48074E-11	1.00341E-08
Case t2 X Case t3	ENSECAG00000015657	TNNT1	1.4281332	1.20509E-10	2.17179E-08
Case t2 X Case t3	ENSECAG00000004763	XIRP1	1.293548826	1.94338E-10	3.44845E-08
Case t2 X Case t3	ENSECAG00000009742	S100A12	2.019360365	2.1197E-10	3.70434E-08
Case t2 X Case t3	ENSECAG00000010089	DUSP6	1.288609136	3.23354E-10	5.56652E-08
Case t2 X Case t3	ENSECAG00000007960	LIMD2	1.60180998	4.69878E-10	7.96997E-08
Case t2 X Case t3	ENSECAG00000019932	LGALS9C	1.413355658	5.05261E-10	8.44592E-08
Case t2 X Case t3	ENSECAG00000000419	TRAC	1.489322423	5.28441E-10	8.63161E-08
Case t2 X Case t3	ENSECAG00000020452	TNNC1	1.369704764	5.31337E-10	8.63161E-08
Case t2 X Case t3	ENSECAG00000010020	HBB	2.119562201	6.78278E-10	1.08656E-07
Case t2 X Case t3	ENSECAG00000014875	PTPRCAP	1.4345552	7.53092E-10	1.18988E-07
Case t2 X Case t3	ENSECAG00000016339	ADAMTS1	1.250351605	8.17222E-10	1.27376E-07
Case t2 X Case t3	ENSECAG00000024788	CORO1A	1.349284846	1.16008E-09	1.78404E-07
Case t2 X Case t3	ENSECAG00000007867	MYL2	1.316652752	1.4061E-09	2.13394E-07
Case t2 X Case t3	ENSECAG00000017410	MYL3	1.309558654	2.60811E-09	3.90675E-07

Case t2 X Case t3	ENSECAG00000018395	TYROBP	1.309718806	5.23159E-09	7.73605E-07
Case t2 X Case t3	ENSECAG00000020463	HSPB1	1.187957489	5.31064E-09	7.75354E-07
Case t2 X Case t3	ENSECAG00000015834	DMD	-1.201494777	5.69262E-09	8.20733E-07
Case t2 X Case t3	ENSECAG00000014552	ENSECAG00000014552	-1.223198618	6.2907E-09	8.95765E-07
Case t2 X Case t3	ENSECAG00000017799	MYO1G	1.308878054	6.98058E-09	9.81879E-07
Case t2 X Case t3	ENSECAG00000009192	SRGN	1.203957895	7.67809E-09	1.06698E-06
Case t2 X Case t3	ENSECAG00000009474	ENSECAG00000009474	1.97331704	8.21324E-09	1.12776E-06
Case t2 X Case t3	ENSECAG00000000763	CSRNP1	1.311913009	8.31742E-09	1.12862E-06
Case t2 X Case t3	ENSECAG00000024888	CCL5	1.28672592	9.00109E-09	1.20719E-06
Case t2 X Case t3	ENSECAG00000016776	TNNI1	1.257187544	9.85638E-09	1.30671E-06
Case t2 X Case t3	ENSECAG00000009363	IRG1	1.854095132	1.0617E-08	1.39155E-06
Case t2 X Case t3	ENSECAG00000011254	SMG1	-1.260155461	1.43074E-08	1.85417E-06
Case t2 X Case t3	ENSECAG00000012148	NMES1	1.690723335	1.50469E-08	1.92834E-06
Case t2 X Case t3	ENSECAG00000002668	RHOB	1.148074246	1.77873E-08	2.2545E-06
Case t2 X Case t3	ENSECAG00000012572	ENSECAG00000012572	-1.183876651	1.93002E-08	2.41966E-06
Case t2 X Case t3	ENSECAG00000020638	NEB	-1.122713663	2.39429E-08	2.96943E-06
Case t2 X Case t3	ENSECAG00000023322	ID3	1.158760487	2.48422E-08	3.04819E-06
Case t2 X Case t3	ENSECAG00000019124	LILRA5	1.516827666	2.63168E-08	3.19514E-06
Case t2 X Case t3	ENSECAG00000001159	FOXS1	1.260488919	2.75603E-08	3.31126E-06
Case t2 X Case t3	ENSECAG00000001784	CD14	1.168974284	3.75389E-08	4.46364E-06
Case t2 X Case t3	ENSECAG00000016508	HK3	1.29768793	3.88245E-08	4.56941E-06
Case t2 X Case t3	ENSECAG00000018028	TLR2	1.357514574	4.04357E-08	4.71096E-06
Case t2 X Case t3	ENSECAG00000010918	SELP	1.305331929	5.07243E-08	5.85055E-06

Table 15. **Comparisons between the effect of training in cases and controls (double difference).** The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl Gene ID	Gene	logFC	PValue	FDR
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000008335	PDK4	-3.493037034	4.96122E-24	2.86114E-20
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000016867	CISH	-3.792464215	3.41662E-24	2.86114E-20
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013728	MFSD2A	-3.934524569	3.58301E-23	1.37755E-19
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000018927	ANGPTL4	-4.347620478	1.58601E-19	4.57327E-16
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000000548	IGHM	3.422578748	2.39987E-19	5.53601E-16
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000014658	SCD	2.792882878	1.70865E-16	3.28459E-13
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013532	CCDC162P	-3.608370261	2.76014E-16	4.54792E-13
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000011811	ENSECAG00000011811	-3.716882066	6.10348E-16	8.7997E-13

Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000003774	IGHG1	3.468734353	1.77083E-15	2.26942E-12
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013692	IGL	4.316407242	3.42804E-15	3.9539E-12
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000022841	NMRK2	2.462906634	4.60312E-15	4.82658E-12
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000014500	ART5	-2.351614959	1.75959E-14	1.69125E-11
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000016486	ARRDC2	-2.412914117	3.37972E-14	2.99859E-11
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000015109	IGL	3.71109855	1.83393E-13	1.5109E-10
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000000816	AMPD3	-2.397086735	5.48344E-13	3.95287E-10
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000020122	SELL	2.628961698	5.15141E-13	3.95287E-10
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000011895	ILT11B	3.001168096	1.49389E-12	1.01356E-09
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000019557	LDLR	2.216682949	2.78922E-12	1.78727E-09
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000005186	HMGB1	5.341137575	4.38147E-12	2.65978E-09
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000001101	SESN1	-2.055316588	1.02591E-11	5.91645E-09
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000018324	SLC43A2	-2.081769899	1.72751E-11	9.48815E-09
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000005487	IGLC7	4.100649508	8.89203E-11	4.66185E-08
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000009075	MPZ	-2.393397776	9.79764E-11	4.9133E-08
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000002376	CHRNG	-2.603185935	1.24787E-10	5.99704E-08
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000012472	COPS4	4.326125448	3.38041E-10	1.55959E-07
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000014861	GRAMD2	2.329533153	1.0268E-09	4.55505E-07
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000018228	IGSF6	2.384168002	1.136E-09	4.85282E-07
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013836	PKLR	-2.776604306	1.74133E-09	7.17304E-07
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000024506	ARMC2	-2.066890128	2.01103E-09	7.99836E-07
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000004107	DDIT4	-1.828966441	3.16194E-09	1.21566E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000019083	SORL1	2.066805753	3.60144E-09	1.33997E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013457	CSF3R	2.178250644	5.21136E-09	1.87837E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000007258	IGHG1	2.880562427	5.46045E-09	1.90851E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000011304	FKBP5	-1.816883506	5.84507E-09	1.98285E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000011258	RFXAP	-1.757145014	1.06718E-08	3.51682E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013291	PER1	-1.749458223	1.21701E-08	3.89915E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000009363	IRG1	2.468727808	1.52993E-08	4.64375E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000011493	ARMC12	-1.798445513	1.52274E-08	4.64375E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000019922	ADAMDEC1	2.596207929	1.87829E-08	5.55492E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000012751	CCKBR	2.231184237	2.04443E-08	5.75133E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000026887	TSPAN8	-1.718399093	2.0049E-08	5.75133E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000023399	GADD45G	-1.815538962	2.1516E-08	5.9087E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000020162	PRR5L	-1.756210925	3.18836E-08	8.55223E-06
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000023728	DDIT4L	-1.661272698	4.57829E-08	1.20014E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000010918	SELP	1.95398925	6.12615E-08	1.53606E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000012395	S100B	-2.277994451	6.0288E-08	1.53606E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000009625	IGHA1	2.540219168	9.40344E-08	2.30764E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000009370	ORC1	-2.073471946	1.31071E-07	3.14952E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000016002	FBXO32	-1.625352808	1.44828E-07	3.40908E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000019552	SLCO1C1	-2.419179259	1.69818E-07	3.91736E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000019795	ARID5B	-1.632961495	1.73508E-07	3.92401E-05

Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000024873	C1orf110	2.111534393	1.78746E-07	3.96472E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000008555	LTB	2.165138463	1.92318E-07	4.18527E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000017758	SH3KBP1	1.543913257	3.65209E-07	7.80058E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000000368	MT1A	-1.859123905	4.02219E-07	8.32345E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000023076	GPCPD1	-1.530815255	4.04121E-07	8.32345E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000006487	METTL21EP	1.894091686	4.27251E-07	8.64546E-05
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000007960	LIMD2	1.903686488	5.49002E-07	0.000107716
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000020809	C3orf18	-1.877145944	5.50999E-07	0.000107716
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013081	MMP9	2.094235107	6.26331E-07	0.000120402
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013302	LRIT3	1.776649067	7.78873E-07	0.000142718
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000017956	TOR3A	1.518978148	7.78471E-07	0.000142718
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000024993	CYR61	-1.595454215	7.79542E-07	0.000142718
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000021110	ENSECAG00000021110	2.188733271	8.61224E-07	0.000155209
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000015010	CYP4F3	2.003100551	8.95462E-07	0.000158896
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000004066	APOLD1	-1.872247484	9.77055E-07	0.000170733
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000016679	GALNT15	-1.605751767	9.91772E-07	0.000170733
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000002437	NFIL3	1.465191139	1.64683E-06	0.000279331
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000001290	CABP1	1.590830173	1.70518E-06	0.000285038
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000018476	TPX2	-1.482187953	1.89903E-06	0.000312905
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000000591	TRAF3IP2	-1.595454765	2.12189E-06	0.000344703
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000011731	HORMAD1	2.020373099	2.31513E-06	0.000370871
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000014255	SAG	1.700565611	2.34887E-06	0.000371121
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000001394	MPEG1	1.835494677	2.46493E-06	0.00038412
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000015730	NAPSA	2.01604271	2.49775E-06	0.00038412
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000021092	KLF15	-1.440983499	2.55012E-06	0.000387014
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000020485	PCOLCE2	-1.520597225	2.68647E-06	0.000402412
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000011075	RFX3	-1.451974041	2.95948E-06	0.000437624
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000023280	FADS1	1.581068934	3.48182E-06	0.000508345
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000019779	KCNS3	1.413131296	3.62593E-06	0.000522769
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000009180	TLR8	2.154214622	3.82728E-06	0.000544986
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000020059	FGF4	-1.948259987	4.10225E-06	0.000577016
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000000400	AMICA1	1.678231149	5.12722E-06	0.000705942
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000010020	HBB	2.756559142	5.14124E-06	0.000705942
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013781	KIAA1211	-1.675443309	5.64101E-06	0.000764004
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000016899	FAM134B	-1.377575494	5.69658E-06	0.000764004
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000006153	C5AR1	1.758422358	6.00536E-06	0.000796158
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000009488	SYK	1.784977723	6.13719E-06	0.000804391
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013745	SNX10	1.819990056	6.82795E-06	0.000884872
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000018029	THSD1	-1.360199365	7.46282E-06	0.000956402
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000000291	PLCG2	1.386435511	1.09675E-05	0.001387472
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000015275	MT2A	-1.374624615	1.11874E-05	0.001387472
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000017436	TREM1	2.081526634	1.11435E-05	0.001387472

Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000013723	SLC7A8	1.33899928	1.16163E-05	0.001425347
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000015500	ENSECAG00000015500	2.09757972	1.23667E-05	0.001501442
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000006674	RRAD	1.318039042	1.28768E-05	0.001547099
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000023457	GPR183	1.941546839	1.35794E-05	0.001614688
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000010108	ITPRIP	-1.396310825	1.59721E-05	0.001865478
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000010185	IFIT2	1.541130915	1.6012E-05	0.001865478
Case t1 - Case t2 X Control t1 - Control t2	ENSECAG00000005489	PTAFR	1.787160092	1.75293E-05	0.002021828

Table 16. **Comparisons between the effect of exercise in cases and controls (double difference).** The list shows the top 100 genes differentially expressed with FDR<0.05

Comparison	Ensembl Gene ID	Gene Name	logFC	PValue	FDR
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000011895	ILT11B	-3.700267044	1.33417E-18	1.53884E-14
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000005487	IGLC7	-4.749828551	3.05813E-17	1.61901E-13
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000020122	SELL	-3.225458636	4.21106E-17	1.61901E-13
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010847	PGLYRP1	-3.721698591	7.22114E-17	2.08222E-13
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000013457	CSF3R	-3.074511081	1.09579E-15	2.52778E-12
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000013081	MMP9	-3.295642061	1.51545E-15	2.9132E-12
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000000400	AMICA1	-2.822624572	6.76733E-14	1.11506E-10
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000017436	TREM1	-3.101915574	2.75996E-13	3.97917E-10
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000019922	ADAMDEC1	-3.262182853	5.96629E-13	7.64614E-10
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000008555	LTB	-2.82910483	2.06382E-12	2.38041E-09
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000003774	IGHG1	-2.859939377	2.80715E-12	2.49059E-09
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000005186	HMGB1	-5.228995048	2.42972E-12	2.49059E-09
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000019932	LGALS9C	-2.507087061	2.59919E-12	2.49059E-09
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009363	IRG1	-3.022610333	5.78044E-12	4.76225E-09
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000019083	SORL1	-2.548273671	8.02132E-12	6.16786E-09
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000018228	IGSF6	-2.688283335	1.35312E-11	9.7543E-09
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000012148	NMES1	-3.002444769	2.05163E-11	1.38379E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000018649	UCP2	-2.386061991	2.15954E-11	1.38379E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000015010	CYP4F3	-2.65826464	2.71763E-11	1.64974E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000004180	HSPA6	-2.705035144	2.87983E-11	1.6608E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000023992	S100P	-2.485513848	4.64183E-11	2.54947E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009625	IGHA1	-2.789680209	5.4331E-11	2.84842E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000004312	IL1RN	-2.723709132	9.29302E-11	4.66025E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010185	IFIT2	-2.430795128	1.11373E-10	5.35238E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000012472	COPS4	-4.495591031	1.19869E-10	5.44165E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000021110	ENSECAG	-2.893511188	1.22666E-10	5.44165E-08

		00000021110			
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000024788	CORO1A	-2.214543452	1.63664E-10	6.99149E-08
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009192	SRGN	-2.068425374	3.68813E-10	1.51925E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000013692	IGL	-3.082838319	4.18814E-10	1.66572E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000000794	EMB	-2.690474342	4.33807E-10	1.66784E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000023216	VNN2	-2.186570963	4.81243E-10	1.79054E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010496	LILRB4	-3.046489353	5.08458E-10	1.83267E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000015500	ENSECAG00000015500	-2.971479994	6.60956E-10	2.31014E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000007960	LIMD2	-2.372634321	7.03829E-10	2.38764E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000014338	EGR1	-2.033276557	8.91445E-10	2.93769E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010918	SELP	-2.253015046	1.13668E-09	3.6418E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000005081	CD3G	-2.378491343	1.23152E-09	3.83903E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000019124	LILRA5	-2.479281806	1.31528E-09	3.94445E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000024888	CCL5	-2.084978314	1.33374E-09	3.94445E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000006561	TRBC2	-2.098882749	1.49699E-09	4.31658E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000013745	SNX10	-2.451216177	1.68906E-09	4.75161E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000016508	HK3	-2.230025104	1.81158E-09	4.97495E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000003816	JUNB	-1.931563897	2.08682E-09	5.59753E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000016721	KIAA0226L	-2.554547498	2.19435E-09	5.75218E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000018395	TYROBP	-2.096004716	3.18461E-09	8.1625E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000019527	SELPLG	-2.334331202	3.54539E-09	8.88968E-07
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000000419	TRAC	-2.125409118	5.38461E-09	1.30268E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010500	ITGA4	-2.310995741	5.42124E-09	1.30268E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009742	S100A12	-2.680913805	5.88207E-09	1.38457E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000024457	SPI1_PU.1	-2.165422882	6.21429E-09	1.43351E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000011890	ARHGAP30	-2.055683961	6.50724E-09	1.47166E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000011467	ALPL	-2.010014322	6.64461E-09	1.47383E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000022135	VNN1	-2.291625358	7.1298E-09	1.55161E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000014635	CD3E	-2.178724514	8.97611E-09	1.91723E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000017799	MYO1G	-1.975760811	9.82747E-09	2.06091E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000011994	CYFIP2	-2.249022665	1.14506E-08	2.35842E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000011733	RNASE6	-2.029589202	1.31128E-08	2.65339E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010020	HBB	-2.703851364	1.37985E-08	2.744E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000007837	FGR	-1.91333279	1.57644E-08	3.08181E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000015730	NAPSA	-2.329142973	1.80821E-08	3.47599E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000018113	LYZ	-1.964649709	1.85813E-08	3.51339E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000014260	FOS	-1.875001557	1.93582E-08	3.60124E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000023457	GPR183	-2.334375263	2.17441E-08	3.9809E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000011917	DOCK8	-2.003423114	2.32637E-08	4.19256E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010263	SASH3	-2.290350755	2.97502E-08	5.27905E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000000436	FCN1	-2.201588713	3.04748E-08	5.3257E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009488	SYK	-2.209233495	4.15293E-08	7.14925E-06

Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000023430	TRPM2	-2.132410191	5.45452E-08	9.25183E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000013534	CD53	-2.013422633	5.76233E-08	9.63228E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000004162	EVI2B	-2.068210539	5.93179E-08	9.77389E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000018339	IKZF1	-2.428211372	6.0624E-08	9.84841E-06
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000018028	TLR2	-2.08182229	6.3883E-08	1.02337E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000012029	TBC1D10C	-2.268495829	6.8951E-08	1.08943E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009282	GZMH	-2.293328115	7.99386E-08	1.24596E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000007258	IGHG1	-2.478861748	9.16454E-08	1.40938E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000017457	CYTIP	-1.932888956	1.38184E-07	2.09713E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000012179	RGS2	-1.788376736	1.51821E-07	2.27416E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010928	MMP25	-2.002613228	1.6158E-07	2.35907E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000021765	TSPO	-1.837750868	1.61044E-07	2.35907E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009012	IER2	-1.658100148	2.04572E-07	2.94942E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000012738	COTL1	-1.860346707	2.23731E-07	3.18582E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000015742	FGD3	-1.966875942	2.74939E-07	3.86725E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009535	FCER1G	-1.783326767	2.96287E-07	4.11732E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000000548	IGHM	-1.852543348	4.01344E-07	5.51084E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000018996	PLEK	-1.885025848	4.80676E-07	6.52249E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000016449	ILT11A	-2.191857588	5.21993E-07	7.00078E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000001141	NAIP	-2.240956853	5.64873E-07	7.40369E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000026984	GPSM3	-1.968535415	5.62072E-07	7.40369E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009075	MPZ	1.842645541	6.13573E-07	7.95163E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000006153	C5AR1	-1.933402662	6.8455E-07	8.67649E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000014875	PTPRCAP	-1.75461438	6.83141E-07	8.67649E-05
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000017644	CFP	-1.808064321	9.92219E-07	0.000124394
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000009958	DUSP1	-1.500559818	1.03776E-06	0.000128705
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000010977	HMHA1	-1.899568828	1.06559E-06	0.000129603
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000016634	RPS2	-1.794909152	1.06748E-06	0.000129603
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000023256	LCK	-1.825708901	1.27648E-06	0.000153364
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000016962	APOBR	-1.880188828	1.32197E-06	0.000157192
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000017707	HCK	-2.023327464	1.65424E-06	0.000192909
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000023596	RASSF2	-1.917177054	1.6558E-06	0.000192909
Case t2 - Case t3 X Control t2 - Control t3	ENSECAG00000015034	PLAC8	-1.731486151	1.80965E-06	0.000208725

Table 17. DAVID clusters based on biological function for all of the comparisons. The list shows the top 10 cluster are represented

CASE X CONTROL t1

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
GOTERM_BP_FAT	GO:0001775~cell activation	39	5.28E-16	371	4.954985584	9.77E-13
GOTERM_BP_FAT	GO:0045321~leukocyte activation	34	1.97E-14	371	5.12298679	3.47E-11
GOTERM_BP_FAT	GO:0006955~immune response	57	1.69E-13	371	3.012211415	2.99E-10
SP_PIR_KEYWORDS	signal	139	2.09E-12	463	1.776816747	2.96E-09
GOTERM_BP_FAT	GO:0002684~positive regulation of immune system process	31	2.54E-12	371	4.749462049	4.47E-09
UP_SEQ_FEATURE	signal peptide	139	2.86E-12	462	1.769368631	4.74E-09
GOTERM_BP_FAT	GO:0046649~lymphocyte activation	28	5.73E-12	371	5.130558453	1.01E-08
GOTERM_BP_FAT	GO:0050865~regulation of cell activation	25	7.09E-11	371	5.209087409	1.25E-07
SP_PIR_KEYWORDS	disulfide bond	124	1.04E-10	463	1.761795582	1.47E-07
GOTERM_BP_FAT	GO:0006952~defense response	48	1.47E-10	371	2.845940438	2.60E-07

CASE X CONTROL t2

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
SP_PIR_KEYWORDS	signal	93	4.16E-12	271	2.031058757	5.77E-09
GOTERM_CC_FAT	GO:0005576~extracellular region	73	5.26E-12	206	2.253504323	6.88E-09
UP_SEQ_FEATURE	signal peptide	93	5.99E-12	271	2.018176554	9.5E-09
SP_PIR_KEYWORDS	Secreted	61	1.24E-11	271	2.563439578	1.72E-08
SP_PIR_KEYWORDS	disulfide bond	83	1.74E-10	271	2.01476141	0.000000242
SP_PIR_KEYWORDS	extracellular matrix	21	2.19E-10	271	6.184792761	0.000000303
GOTERM_CC_FAT	GO:0044421~extracellular region part	44	4.89E-10	206	2.843891586	0.000000639
UP_SEQ_FEATURE	disulfide bond	79	1.5E-09	271	1.976476178	0.00000239
GOTERM_CC_FAT	GO:0005578~proteinaceous extracellular matrix	24	1.73E-09	206	4.653640777	0.00000226

GOTERM_CC_FAT	GO:0031012~extracellular matrix	24	7.27E-09	206	4.316420431	0.00000951

CASE X CONTROL t3

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
GOTERM_BP_FAT	GO:0009991~response to extracellular stimulus	29	1.17972E-07	561	3.178674445	0.000212246
GOTERM_BP_FAT	GO:0010033~response to organic substance	61	1.65446E-07	561	2.040165051	0.000297658
UP_SEQ_FEATURE	sequence variant	512	3.40794E-07	716	1.139709527	0.000591765
GOTERM_BP_FAT	GO:0010035~response to inorganic substance	27	3.59908E-07	561	3.176001043	0.00064752
GOTERM_CC_FAT	GO:0009986~cell surface	36	4.58914E-07	511	2.587623996	0.000636047
GOTERM_BP_FAT	GO:0006955~immune response	57	9.88464E-07	561	1.992032861	0.00177836
SP_PIR_KEYWORDS	phosphoprotein	333	1.01098E-06	717	1.229988991	0.001463873
SP_PIR_KEYWORDS	polymorphism	491	1.34079E-06	717	1.140440266	0.001941425
SP_PIR_KEYWORDS	signal	170	1.76868E-06	717	1.403261453	0.002560984
SP_PIR_KEYWORDS	actin-binding	27	1.93454E-06	717	2.932512324	0.002801142

CASE t1 X CASE t2

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
GOTERM_BP_FAT	GO:0014866~skeletal myofibril assembly	3	0.000510136	81	83.50617284	0.815056651
GOTERM_BP_FAT	GO:0045471~response to ethanol	5	0.000550345	81	13.04783951	0.879033544
GOTERM_BP_FAT	GO:0001666~response to hypoxia	6	0.001173746	81	7.478164732	1.866000879
GOTERM_BP_FAT	GO:0070482~response to oxygen levels	6	0.001471828	81	7.106908327	2.334661762
UP_SEQ_FEATURE	domain:LDL-receptor class A 6	3	0.001474173	102	51.10427807	2.079536352
UP_SEQ_FEATURE	domain:LDL-receptor class A 7	3	0.001474173	102	51.10427807	2.079536352
UP_SEQ_FEATURE	domain:LDL-receptor class A 5	3	0.00176292	102	46.84558824	2.482117355
GOTERM_BP_FAT	GO:0002714~positive regulation of B cell mediated immunity	3	0.001834911	81	45.54882155	2.902683904
GOTERM_BP_FAT	GO:0002891~positive regulation of immunoglobulin mediated immune response	3	0.001834911	81	45.54882155	2.902683904
UP_SEQ_FEATURE	domain:LDL-receptor class A 4	3	0.00316146	102	35.13419118	4.410322607

CASE t2 X CASE t3

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR

SP_PIR_KEYWORDS	phosphoprotein	1294	6.63E-68	2390	1.433879058	1.00E-64
GOTERM_BP_FAT	GO:0006414~translational elongation	67	4.46E-33	1855	4.837746524	8.31E-30
GOTERM_CC_FAT	GO:0033279~ribosomal subunit	71	2.88E-29	1680	4.220247396	4.32E-26
SP_PIR_KEYWORDS	acetylation	513	3.47E-29	2390	1.566863036	5.25E-26
SP_PIR_KEYWORDS	ribosomal protein	86	4.56E-29	2390	3.681585507	6.89E-26
SP_PIR_KEYWORDS	ribosome	51	9.22E-29	2390	5.62265719	1.39E-25
GOTERM_MF_FAT	GO:0003735~structural constituent of ribosome	81	1.16E-26	1776	3.524583736	1.91E-23
KEGG_PATHWAY	hsa03010:Ribosome	57	2.05E-26	754	4.418503613	2.53E-23
GOTERM_CC_FAT	GO:0022626~cytosolic ribosome	52	1.30E-25	1680	4.88436214	1.95E-22
SP_PIR_KEYWORDS	cytoplasm	599	5.46E-25	2390	1.446825383	8.25E-22
GOTERM_CC_FAT	GO:0005739~mitochondrion	222	2.75E-12	1680	1.553863845	4.11E-09
GOTERM_CC_FAT	GO:0015629~actin cytoskeleton	76	7.29E-11	1680	2.149566295	1.09E-07

Double difference t1 X t2

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
SP_PIR_KEYWORDS	transmembrane protein	26	2.90E-08	208	3.745132399	3.94E-05
GOTERM_CC_FAT	GO:0005887~integral to plasma membrane	35	1.04E-07	140	2.689814815	1.30E-04
GOTERM_CC_FAT	GO:0031226~intrinsic to plasma membrane	35	1.79E-07	140	2.630041152	2.23E-04
SP_PIR_KEYWORDS	disulfide bond	61	2.89E-07	208	1.929218076	3.94E-04
SP_PIR_KEYWORDS	signal	65	4.39E-07	208	1.849519231	5.97E-04
UP_SEQ_FEATURE	signal peptide	65	4.56E-07	207	1.846666667	6.96E-04
GOTERM_BP_FAT	GO:0007159~leukocyte adhesion	7	9.07E-07	166	20.37349398	0.001521206
GOTERM_BP_FAT	GO:0007229~integrin-mediated signaling pathway	9	2.05E-06	166	10.4777969	0.003443328
KEGG_PATHWAY	hsa04514:Cell adhesion molecules (CAMs)	12	2.15E-06	74	6.246928747	0.002304074
UP_SEQ_FEATURE	disulfide bond	57	2.47E-06	207	1.866974104	0.003766139

Double difference t2 X t3

Category	Term	Count	PValue	List Total	Fold Enrichment	FDR
GOTERM_BP_FAT	GO:0001775~cell activation	33	5.01E-21	178	8.738675958	8.34E-18
GOTERM_BP_FAT	GO:0045321~leukocyte activation	30	5.51E-20	178	9.421487603	9.17E-17

GOTERM_BP_FAT	GO:0006955~immune response	46	6.78E-20	178	5.066666667	1.13E-16
GOTERM_BP_FAT	GO:0006952~defense response	39	6.08E-16	178	4.819512195	9.21E-13
GOTERM_BP_FAT	GO:0046649~lymphocyte activation	23	1.36E-14	178	8.783919598	2.25E-11
GOTERM_BP_FAT	GO:0042110~T cell activation	19	4.35E-14	178	11.46031746	7.22E-11
GOTERM_BP_FAT	GO:0002684~positive regulation of immune system process	23	5.53E-13	178	7.344537815	9.20E-10
GOTERM_BP_FAT	GO:0050867~positive regulation of cell activation	16	1.42E-11	178	10.95495495	2.36E-08
GOTERM_BP_FAT	GO:0050865~regulation of cell activation	18	1.27E-10	178	7.817142857	2.11E-07
GOTERM_CC_FAT	GO:0044459~plasma membrane part	62	2.11E-10	161	2.234344471	2.69E-07

Table 18. Summary of pathways used by GSEA and their enrichment scores and FDRs

COMPARISON CASE X CONTROL T2						
PATHWAY	SIZE	ES	NES	NOM p-val	FDR q-val	
FATTY_ACID_METABOLIC_PROCESS	39	-0.5141444	-1.584686	0.011320755	0.07342478	**ENRICHED CONTROLS**
COMPARISON CASE X CONTROL T2						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	101	-0.4952191	-1.4742914	0.14604463	0.23013428	**ENRICHED CONTROLS**
KEGG_OXIDATIVE_PHOSPHORYLATION	90	-0.4816668	-1.4316875	0.19631901	0.21576932	**ENRICHED CONTROLS**
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	54	-0.5169294	-1.3437802	0.22129436	0.20644842	**ENRICHED CONTROLS**
COMPARISON CASE X CONTROL T2						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
MITOCHONDRIAL_RESPIRATORY_CHAIN	21	-0.52548486	-1.3155179	0.21713148	0.2634178	**ENRICHED CONTROLS**
MITOCHONDRIAL_MEMBRANE	68	-0.42892662	-1.3052319	0.25	0.22932892	**ENRICHED CONTROLS**
COMPARISON CASE X CONTROL T1						
PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	35	0.64731264	2.0261514	0	0.009430527	**ENRICHED CASES**
REACTOME_GLYCOLYSIS	21	0.58073455	1.8121207	0.031189084	0.094652504	**ENRICHED CASES**

COMPARISON CASE T1 X CASE T2

PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
MOOTHA_GLYCOGEN_METABOLISM	16	0.6143548	1.4592139	0.062248997	0.11375212	**ENRICHED CASES T1**

COMPARISON CASE X CONTROL T1

PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
MOOTHA_GLYCOGEN_METABOLISM	16	0.4889237	1.3183392	0.16302186	0.22617124	**ENRICHED CASES**

COMPARISON CASE X CONTROL T1

PATHWAYS	SIZE	ES	NES	NOM p-val	FDR q-val	
REGULATION_OF_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	25	0.4370639	1.4382352	0.036363635	0.21895188	**ENRICHED CASES**
NEGATIVE_REGULATION_OF_CELLULAR_COMPONENT_ORGANIZATION_AND_BIOGENESIS	24	0.47610074	1.4318788	0.060796645	0.17171904	**ENRICHED CASES**
CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	156	0.38595834	1.361863	0.10309278	0.22401287	**ENRICHED CASES**
ACTIN_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	83	0.40572113	1.2894213	0.1764706	0.24646045	**ENRICHED CASES**

Table 19. **Summary of pathways used by GSEA and the genes from the expression data set that contributed to the enrichment.** ES: enrichment score. Core enrichment YES: Genes that contributed for the pathway to be significant

FATTY ACID METABOLIC PROCESS CASE X CONTROL t2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
ALOX12	arachidonate 12-lipoxygenase	-0.001875266	No
ALOX15B	arachidonate 15-lipoxygenase, type B	-0.021472909	No
CPT1B	carnitine palmitoyltransferase 1B (muscle)	-0.01708801	No
PTGES	prostaglandin E synthase	-0.024569668	No
CYP2J2	cytochrome P450, family 2, subfamily J, polypeptide 2	-0.03541219	No
CYP4F3	cytochrome P450, family 4, subfamily F, polypeptide 3	-0.14990844	No
SLC27A5	solute carrier family 27 (fatty acid transporter), member 5	-0.15139365	No
BRCA1	breast cancer 1, early onset	-0.15417928	No
CPT1A	carnitine palmitoyltransferase 1A (liver)	-0.15676837	No
MCAT	malonyl CoA:ACP acyltransferase (mitochondrial)	-0.1827632	No
BDH2	3-hydroxybutyrate dehydrogenase, type 2	-0.19757262	No
PPARA	peroxisome proliferative activated receptor, alpha	-0.1946696	No
PPARD	peroxisome proliferative activated receptor, delta	-0.24661613	No
ADIPOR1	adiponectin receptor 1	-0.2491975	No
MIF	macrophage migration inhibitory factor (glycosylation-inhibiting factor)	-0.24033439	No
ACADVL	acyl-Coenzyme A dehydrogenase, very long chain	-0.3069712	No
ECHS1	enoyl Coenzyme A hydratase, short chain, 1, mitochondrial	-0.3094807	No
ACOT9	acyl-CoA thioesterase 9	-0.29732838	No
HACL1	2-hydroxyacyl-CoA lyase 1	-0.35747975	No
GNPAT	glyceronephosphate O-acyltransferase	-0.33733857	No
ACADM	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	-0.419112	No
CROT	carnitine O-octanoyltransferase	-0.43554103	No
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A	-0.48220152	Yes
ADIPOR2	adiponectin receptor 2	-0.4722387	Yes
LTA4H	leukotriene A4 hydrolase	-0.4513995	Yes
PNPLA8	patatin-like phospholipase domain containing 8	-0.43264836	Yes
ACOX3	acyl-Coenzyme A oxidase 3, pristanoyl	-0.41427162	Yes
DEGS1	degenerative spermatocyte homolog 1, lipid desaturase (Drosophila)	-0.3820845	Yes
HADHB	hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-	-0.35460165	Yes

	Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit		
PPARGC1A	peroxisome proliferative activated receptor, gamma, coactivator 1, alpha	-0.31814194	Yes
ACOT4	acyl-CoA thioesterase 4	-0.30559286	Yes
PTGES3	prostaglandin E synthase 3 (cytosolic)	-0.26580158	Yes
CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	-0.22571754	Yes
ECH1	enoyl Coenzyme A hydratase 1, peroxisomal	-0.19469436	Yes
ALDH5A1	aldehyde dehydrogenase 5 family, member A1 (succinate-semialdehyde dehydrogenase)	-0.16179325	Yes
FADS2	fatty acid desaturase 2	-0.11499819	Yes
PTGS1	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	-0.069887646	Yes
OXSM	3-oxoacyl-ACP synthase, mitochondrial	-0.028105699	Yes
FADS1	fatty acid desaturase 1	0.013948627	Yes

REACTOME TCA CYCLE AND RESPIRATORY ELECTRON TRANSPORT			
CASE X CONTROL t2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
UCP3	uncoupling protein 3 (mitochondrial, proton carrier)	-0.077779755	No
PDPR	-	-0.08084049	No
ADHFE1	alcohol dehydrogenase, iron containing, 1	-0.100758	No
PDK4	pyruvate dehydrogenase kinase, isozyme 4	-0.11849339	No
PDK1	pyruvate dehydrogenase kinase, isozyme 1	-0.13074146	No
UCP2	uncoupling protein 2 (mitochondrial, proton carrier)	-0.19924106	No
PDK2	pyruvate dehydrogenase kinase, isozyme 2	-0.2427115	No
LDHB	lactate dehydrogenase B	-0.27059102	No
ETFB	electron-transfer-flavoprotein, beta polypeptide	-0.2748912	No
NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	-0.27950856	No
ND1	mitochondrially encoded NADH dehydrogenase 1	-0.3084035	No
IDH1	isocitrate dehydrogenase 1 (NADP+), soluble	-0.3174611	No
ATP6	mitochondrially encoded ATP synthase 6	-0.32303014	No
SLC16A1	solute carrier family 16, member 1 (monocarboxylic acid transporter 1)	-0.3297972	No
NDUFA4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4,	-0.36693737	No

	9kDa		
ATP5F1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit B1	-0.37451062	No
ATP5D	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit	-0.40883926	No
IDH2	isocitrate dehydrogenase 2 (NADP+), mitochondrial	-0.4318312	No
SUCLG2	succinate-CoA ligase, GDP-forming, beta subunit	-0.42755577	No
NDUFB10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa	-0.43041644	No
NDUFB7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18kDa	-0.4390531	No
COX5B	cytochrome c oxidase subunit Vb	-0.43678913	No
IDH3B	isocitrate dehydrogenase 3 (NAD+) beta	-0.44030583	No
L2HGDH	L-2-hydroxyglutarate dehydrogenase	-0.4794819	No
UQCRC1	ubiquinol-cytochrome c reductase core protein I	-0.47791737	No
NDUFA12	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 12	-0.47513583	No
NDUFB1	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	-0.47571597	No
SUCLG1	succinate-CoA ligase, GDP-forming, alpha subunit	-0.47174224	No
NDUFA2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	-0.46660233	No
ND6	mitochondrially encoded NADH dehydrogenase 6	-0.48280084	No
SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	-0.48646623	No
COX4I1	cytochrome c oxidase subunit IV isoform 1	-0.4807532	No
ETFDH	electron-transferring-flavoprotein dehydrogenase	-0.48745328	Yes
COX5A	cytochrome c oxidase subunit Va	-0.47994524	Yes
NDUFS5	NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa (NADH-coenzyme Q reductase)	-0.47252077	Yes
ATP5A1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	-0.4678567	Yes
NDUFA7	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 7, 14.5kDa	-0.4625415	Yes
UQCRC2	ubiquinol-cytochrome c reductase core protein II	-0.46068302	Yes
ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	-0.4593392	Yes
NDUFAB1	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	-0.45755434	Yes
BSG	basigin (Ok blood group)	-0.44950238	Yes
ND5	mitochondrially encoded NADH dehydrogenase 5	-0.44753554	Yes
SDHC	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa	-0.43956238	Yes
NDUFB2	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2,	-0.43862152	Yes

	8kDa		
PDK3	pyruvate dehydrogenase kinase, isozyme 3	-0.43044227	Yes
NDUFB9	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	-0.435454	Yes
NDUFA3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3, 9kDa	-0.4412764	Yes
COX3	mitochondrially encoded cytochrome c oxidase III	-0.43556845	Yes
COX6A1	cytochrome c oxidase subunit VIa polypeptide 1	-0.43302214	Yes
CYC1	cytochrome c-1	-0.4257684	Yes
SUCLA2	succinate-CoA ligase, ADP-forming, beta subunit	-0.41726094	Yes
D2HGDH	D-2-hydroxyglutarate dehydrogenase	-0.42121458	Yes
ACO2	aconitase 2, mitochondrial	-0.42224848	Yes
NDUFB5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa	-0.42209738	Yes
COX2	mitochondrially encoded cytochrome c oxidase II	-0.42619342	Yes
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	-0.4242266	Yes
DLD	dihydrolipoamide dehydrogenase	-0.42364955	Yes
NDUFA6	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	-0.41777354	Yes
PDHX	pyruvate dehydrogenase complex, component X	-0.41212803	Yes
OGDH	oxoglutarate (alpha-ketoglutarate) dehydrogenase (lipoamide)	-0.40362877	Yes
NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa (NADH-coenzyme Q reductase)	-0.39775926	Yes
NNT	nicotinamide nucleotide transhydrogenase	-0.39500913	Yes
DLAT	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	-0.3890362	Yes
NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	-0.38227606	Yes
IDH3G	isocitrate dehydrogenase 3 (NAD+) gamma	-0.38258222	Yes
ATP5H	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	-0.37176874	Yes
PDHA1	pyruvate dehydrogenase (lipoamide) alpha 1	-0.3689542	Yes
NDUFA10	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 10, 42kDa	-0.38033962	Yes
MDH2	malate dehydrogenase 2, NAD (mitochondrial)	-0.37475726	Yes
IDH3A	isocitrate dehydrogenase 3 (NAD+) alpha	-0.39041668	Yes
ATP5J	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6	-0.38110152	Yes
NDUFV3	NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa	-0.3685029	Yes
ATP8	mitochondrially encoded ATP synthase 8	-0.35860977	Yes
NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1,	-0.3462651	Yes

	7.5kDa		
COX6B1	cytochrome c oxidase subunit Vib polypeptide 1 (ubiquitous)	-0.338455	Yes
UQCRB	ubiquinol-cytochrome c reductase binding protein	-0.32693925	Yes
PDHB	pyruvate dehydrogenase (lipoamide) beta	-0.31708035	Yes
DLST	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)	-0.30652368	Yes
NDUFS3	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	-0.29791242	Yes
NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49kDa (NADH-coenzyme Q reductase)	-0.28541273	Yes
UQCRH	ubiquinol-cytochrome c reductase hinge protein	-0.27418014	Yes
UQCRFS1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	-0.26170817	Yes
ATP5I	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit E	-0.2478872	Yes
CS	citrate synthase	-0.23318207	Yes
NDUFB4	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	-0.22114788	Yes
NDUFS6	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase)	-0.20644939	Yes
NDUFV2	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	-0.1948261	Yes
NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	-0.18074577	Yes
FH	fumarate hydratase	-0.16779904	Yes
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	-0.15359361	Yes
PDP2	-	-0.14403482	Yes
SDHB	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	-0.14627676	Yes
NDUFV1	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kDa	-0.1289687	Yes
ATP5G1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C1 (subunit 9)	-0.11278582	Yes
CYCS	cytochrome c, somatic	-0.09721736	Yes
COX7A2L	cytochrome c oxidase subunit VIIa polypeptide 2 like	-0.08111649	Yes
NDUFB3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	-0.06670696	Yes
NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	-0.048332803	Yes
ETFA	electron-transfer-flavoprotein, alpha polypeptide (glutaric aciduria II)	-0.0320098	Yes
LDHA	lactate dehydrogenase A	-0.016022952	Yes
ATP5J2	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F2	0.004851622	Yes

KEEG OXIDATIVE PHOSPHORILATION CASE X CONTROL t2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PPA2	pyrophosphatase (inorganic) 2	-0.1383572	No
ATP6V1D	ATPase, H+ transporting, lysosomal 34kDa, V1 subunit D	-0.15502721	No
COX10	COX10 homolog, cytochrome c oxidase assembly protein, heme A: farnesyltransferase (yeast)	-0.18437451	No
ATP6V1A	ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A	-0.2146047	No
COX15	COX15 homolog, cytochrome c oxidase assembly protein	-0.2532385	No
NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	-0.28776148	No
ATP6V1B2	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B2	-0.28587264	No
ND1	mitochondrially encoded NADH dehydrogenase 1	-0.31285247	No
ATP6	mitochondrially encoded ATP synthase 6	-0.33049637	No
ATP6V0A1	ATPase, H+ transporting, lysosomal V0 subunit a1	-0.338963	No
ATP6V1G1	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G1	-0.34955636	No
TCIRG1	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 subunit A3	-0.36079842	No
COX4I2	cytochrome c oxidase subunit IV isoform 2 (lung)	-0.3595485	No
ATP6V0B	ATPase, H+ transporting, lysosomal 21kDa, V0 subunit b	-0.35486048	No
NDUFA4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa	-0.3537128	No
ATP5F1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit B1	-0.36062658	No
ATP5D	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit	-0.39418468	No
PPA1	pyrophosphatase (inorganic) 1	-0.39780587	No
ATP6V1C1	ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C1	-0.40256605	No
ATP6V0A2	ATPase, H+ transporting, lysosomal V0 subunit a2	-0.41142514	No
NDUFB10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa	-0.40730393	No
NDUFB7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18kDa	-0.41508955	No
COX5B	cytochrome c oxidase subunit Vb	-0.41197765	No
ATP6V0E2	ATPase, H+ transporting V0 subunit e2	-0.4440703	No
UQCRC1	ubiquinol-cytochrome c reductase core protein I	-0.45747873	No
NDUFB1	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	-0.46119997	No
NDUFA2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	-0.45813844	No
ND6	mitochondrially encoded NADH dehydrogenase 6	-0.47325832	Yes
ATP6V1E1	ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E1	-0.47262648	Yes

SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	-0.46717972	Yes
COX4I1	cytochrome c oxidase subunit IV isoform 1	-0.46037856	Yes
COX5A	cytochrome c oxidase subunit Va	-0.46629417	Yes
NDUFS5	NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa (NADH-coenzyme Q reductase)	-0.45775256	Yes
ATP5A1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	-0.45195973	Yes
NDUFA7	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 7, 14.5kDa	-0.4455115	Yes
UQCRC2	ubiquinol-cytochrome c reductase core protein II	-0.4425063	Yes
ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	-0.4400004	Yes
NDUFAB1	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	-0.43704113	Yes
ND5	mitochondrially encoded NADH dehydrogenase 5	-0.43406275	Yes
SDHC	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa	-0.42490643	Yes
NDUFB2	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	-0.42275935	Yes
NDUFB9	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	-0.42679137	Yes
NDUFA3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3, 9kDa	-0.43133307	Yes
COX3	mitochondrially encoded cytochrome c oxidase III	-0.42434645	Yes
COX6A2	cytochrome c oxidase subunit VIa polypeptide 2	-0.4184304	Yes
COX6A1	cytochrome c oxidase subunit VIa polypeptide 1	-0.4101815	Yes
CYC1	cytochrome c-1	-0.40163118	Yes
ATP6V0E1	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e1	-0.4023538	Yes
NDUFB5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa	-0.41394666	Yes
ATP5G3	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C3 (subunit 9)	-0.40557447	Yes
COX2	mitochondrially encoded cytochrome c oxidase II	-0.40551218	Yes
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	-0.40210256	Yes
NDUFA6	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	-0.40443856	Yes
NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa (NADH-coenzyme Q reductase)	-0.40392387	Yes
ATP6V0D1	ATPase, H+ transporting, lysosomal 38kDa, V0 subunit d1	-0.4020192	Yes
NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	-0.39615637	Yes
ATP5G2	ATP synthase, H+ transporting, mitochondrial F0 complex,	-0.39425188	Yes

	subunit C2 (subunit 9)		
ATP5H	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	-0.3824515	Yes
ATP6V1H	ATPase, H+ transporting, lysosomal 50/57kDa, V1 subunit H	-0.3785939	Yes
ATP6AP1	ATPase, H+ transporting, lysosomal accessory protein 1	-0.37433064	Yes
NDUFA10	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 10, 42kDa	-0.37419033	Yes
COX7A1	cytochrome c oxidase subunit VIIa polypeptide 1 (muscle)	-0.37512848	Yes
ATP6V0C	ATPase, H+ transporting, lysosomal 16kDa, V0 subunit c	-0.37813464	Yes
ATP5J	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6	-0.36921194	Yes
NDUFV3	NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa	-0.35471496	Yes
ATP8	mitochondrially encoded ATP synthase 8	-0.34290394	Yes
NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	-0.32863364	Yes
COX6B1	cytochrome c oxidase subunit Vib polypeptide 1 (ubiquitous)	-0.31886044	Yes
UQCRB	ubiquinol-cytochrome c reductase binding protein	-0.30537233	Yes
NDUFA4L2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2	-0.29027393	Yes
COX11	COX11 homolog, cytochrome c oxidase assembly protein (yeast)	-0.2863223	Yes
NDUFS3	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	-0.27018216	Yes
NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49kDa (NADH-coenzyme Q reductase)	-0.25560707	Yes
UQCRH	ubiquinol-cytochrome c reductase hinge protein	-0.24227697	Yes
UQCRFS1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	-0.22769533	Yes
ATP5I	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit E	-0.21176222	Yes
NDUFB4	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	-0.19768006	Yes
NDUFS6	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase)	-0.18084544	Yes
NDUFV2	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	-0.1670558	Yes
NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	-0.1508006	Yes
COX7A2	cytochrome c oxidase subunit VIIa polypeptide 2 (liver)	-0.13513853	Yes
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	-0.11933161	Yes
SDHB	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	-0.12545924	Yes
NDUFV1	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kDa	-0.105614856	Yes

ATP5G1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C1 (subunit 9)	-0.086865924	Yes
COX7A2L	cytochrome c oxidase subunit VIIa polypeptide 2 like	-0.07075241	Yes
NDUFB3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	-0.053585187	Yes
NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	-0.032419592	Yes
LHPP	-	-0.012746587	Yes
ATP5J2	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F2	0.004846758	Yes

REACTOME RESPIRATORY ELECTRON TRANSPORT CASE X CONTROL t2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
ETFB	electron-transfer-flavoprotein, beta polypeptide	-0.2840789	No
NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	-0.28658172	No
ND1	mitochondrially encoded NADH dehydrogenase 1	-0.31288022	No
NDUFA4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa	-0.37895304	No
NDUFB10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa	-0.4588479	No
NDUFB7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18kDa	-0.46307394	No
COX5B	cytochrome c oxidase subunit Vb	-0.45640588	No
UQCRC1	ubiquinol-cytochrome c reductase core protein I	-0.5051085	Yes
NDUFA12	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 12	-0.49720556	Yes
NDUFB1	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	-0.49259153	Yes
NDUFA2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2,	-0.48529938	Yes

	8kDa		
ND6	mitochondrially encoded NADH dehydrogenase 6	-0.49591264	Yes
SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	-0.49390626	Yes
COX4I1	cytochrome c oxidase subunit IV isoform 1	-0.48253882	Yes
ETFDH	electron-transferring-flavoprotein dehydrogenase	-0.48338273	Yes
COX5A	cytochrome c oxidase subunit Va	-0.4700721	Yes
NDUFS5	NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa (NADH-coenzyme Q reductase)	-0.45684117	Yes
NDUFA7	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 7, 14.5kDa	-0.44888905	Yes
UQCRC2	ubiquinol-cytochrome c reductase core protein II	-0.44107547	Yes
NDUFAB1	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	-0.4399557	Yes
ND5	mitochondrially encoded NADH dehydrogenase 5	-0.4319962	Yes
SDHC	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa	-0.41787323	Yes
NDUFB2	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	-0.4106692	Yes
NDUFB9	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	-0.40949014	Yes
NDUFA3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3, 9kDa	-0.40866888	Yes
COX3	mitochondrially encoded cytochrome c oxidase III	-0.39631727	Yes
COX6A1	cytochrome c oxidase subunit VIa polypeptide 1	-0.38703766	Yes
CYC1	cytochrome c-1	-0.3730457	Yes
NDUFB5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa	-0.3901185	Yes
COX2	mitochondrially encoded cytochrome c oxidase II	-0.38681197	Yes
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	-0.37735277	Yes
NDUFA6	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	-0.37344348	Yes
NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa (NADH-coenzyme Q reductase)	-0.36653602	Yes
NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	-0.36453146	Yes
NDUFA10	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 10, 42kDa	-0.38742667	Yes
NDUFV3	NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa	-0.40431452	Yes
NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	-0.3854751	Yes
COX6B1	cytochrome c oxidase subunit Vib polypeptide 1 (ubiquitous)	-0.3674662	Yes

UQCRB	ubiquinol-cytochrome c reductase binding protein	-0.3457001	Yes
NDUFS3	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	-0.33408844	Yes
NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49kDa (NADH-coenzyme Q reductase)	-0.31080255	Yes
UQCRH	ubiquinol-cytochrome c reductase hinge protein	-0.2886703	Yes
UQCRFS1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	-0.2652342	Yes
NDUFB4	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	-0.24315336	Yes
NDUFS6	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase)	-0.21735151	Yes
NDUFV2	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	-0.19447073	Yes
NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	-0.16908643	Yes
SDHB	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	-0.16829608	Yes
NDUFV1	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kDa	-0.13780451	Yes
CYCS	cytochrome c, somatic	-0.11045275	Yes
COX7A2L	cytochrome c oxidase subunit VIIa polypeptide 2 like	-0.08054676	Yes
NDUFB3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	-0.051807072	Yes
NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	-0.018924028	Yes
ETFA	electron-transfer-flavoprotein, alpha polypeptide (glutaric aciduria II)	0.012561537	Yes

MITOCHONDRIAL RESPIRATORY CHAIN CASE X CONTROL t2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
COX15	COX15 homolog, cytochrome c oxidase assembly protein (yeast)	-0.2475849	No
ND1	mitochondrially encoded NADH dehydrogenase 1	-0.30186266	No
OXA1L	oxidase (cytochrome c) assembly 1-like	-0.32289225	No
UQCRC1	ubiquinol-cytochrome c reductase core protein I	-0.4941741	Yes
NDUFA2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	-0.47838348	Yes
SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	-0.47811702	Yes
SURF1	surfeit 1	-0.46947172	Yes
NDUFAB1	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	-0.44593906	Yes
BCS1L	BCS1-like (yeast)	-0.48089308	Yes

NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	-0.46584612	Yes
NDUFA6	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	-0.4322777	Yes
NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa (NADH-coenzyme Q reductase)	-0.39497527	Yes
NNT	nicotinamide nucleotide transhydrogenase	-0.35341427	Yes
NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	-0.31198752	Yes
NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	-0.33827668	Yes
UQCRB	ubiquinol-cytochrome c reductase binding protein	-0.2828647	Yes
NDUFS3	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	-0.22991017	Yes
NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49kDa (NADH-coenzyme Q reductase)	-0.16503948	Yes
UQCRH	ubiquinol-cytochrome c reductase hinge protein	-0.100903094	Yes
NDUFV1	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kDa	-0.06063692	Yes
NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	0.016554177	Yes

MITOCHONDRIAL MEMBRANE CASE X CONTROL t2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
VDAC3	voltage-dependent anion channel 3	0.023018867	No
BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3	-0.007206425	No
UCP3	uncoupling protein 3 (mitochondrial, proton carrier)	-0.044556625	No
MCL1	myeloid cell leukemia sequence 1 (BCL2-related)	-0.07062096	No
FIS1	fission 1 (mitochondrial outer membrane) homolog (S. cerevisiae)	-0.14470601	No
COX18	COX18 cytochrome c oxidase assembly homolog (S. cerevisiae)	-0.18328793	No
PHB2	prohibitin 2	-0.21242206	No
COX15	COX15 homolog, cytochrome c oxidase assembly protein (yeast)	-0.21088022	No
PPOX	protoporphyrinogen oxidase	-0.26368093	No
BCL2	B-cell CLL/lymphoma 2	-0.2684764	No
ND1	mitochondrially encoded NADH dehydrogenase 1	-0.26528937	No
TIMM13	translocase of inner mitochondrial membrane 13 homolog (yeast)	-0.2802119	No
OXA1L	oxidase (cytochrome c) assembly 1-like	-0.2924411	No
RHOT2	ras homolog gene family, member T2	-0.31599358	No
TIMM10	translocase of inner mitochondrial membrane 10 homolog (yeast)	-0.3386818	No

ATP5D	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit	-0.3562254	No
HCCS	holocytochrome c synthase (cytochrome c heme-lyase)	-0.36811918	No
PMPCA	peptidase (mitochondrial processing) alpha	-0.3885494	No
TIMM50	translocase of inner mitochondrial membrane 50 homolog (S. cerevisiae)	-0.3837079	No
MFN2	mitofusin 2	-0.38769987	No
TIMM8B	translocase of inner mitochondrial membrane 8 homolog B (yeast)	-0.39885062	No
MSTO1	misato homolog 1 (Drosophila)	-0.40797	No
UQCRC1	ubiquinol-cytochrome c reductase core protein I	-0.4052011	No
ABCB7	ATP-binding cassette, sub-family B (MDR/TAP), member 7	-0.39853916	No
NDUFA2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	-0.39990288	No
TIMM9	translocase of inner mitochondrial membrane 9 homolog (yeast)	-0.41715932	Yes
SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	-0.4108912	Yes
MTX2	metaxin 2	-0.40036273	Yes
ATP5A1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	-0.40689746	Yes
SLC25A11	solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier), member 11	-0.39935133	Yes
SURF1	surfeit 1	-0.38947535	Yes
TIMM17B	translocase of inner mitochondrial membrane 17 homolog B (yeast)	-0.37909773	Yes
ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	-0.37184763	Yes
NDUFAB1	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	-0.36533776	Yes
SLC25A3	solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 3	-0.3541259	Yes
OPA1	optic atrophy 1 (autosomal dominant)	-0.37710023	Yes
SLC25A1	solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1	-0.36799163	Yes
VDAC1	voltage-dependent anion channel 1	-0.36720273	Yes
SLC25A15	solute carrier family 25 (mitochondrial carrier; ornithine transporter) member 15	-0.36253235	Yes
BCS1L	BCS1-like (yeast)	-0.35946345	Yes
ATP5G3	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C3 (subunit 9)	-0.35621962	Yes
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	-0.35970587	Yes
VDAC2	voltage-dependent anion channel 2	-0.3515104	Yes

NDUFA6	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	-0.34142992	Yes
OGDH	oxoglutarate (alpha-ketoglutarate) dehydrogenase (lipoamide)	-0.33172283	Yes
NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa (NADH-coenzyme Q reductase)	-0.31973976	Yes
NNT	nicotinamide nucleotide transhydrogenase	-0.3107831	Yes
CASP7	caspase 7, apoptosis-related cysteine peptidase	-0.29399756	Yes
NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	-0.28566563	Yes
ATP5G2	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C2 (subunit 9)	-0.27889383	Yes
RHOT1	ras homolog gene family, member T1	-0.28263167	Yes
IMMT	inner membrane protein, mitochondrial (mitofilin)	-0.28474042	Yes
TOMM22	translocase of outer mitochondrial membrane 22 homolog (yeast)	-0.28202373	Yes
ATP5J	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6	-0.27192336	Yes
NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	-0.25604352	Yes
MPV17	MpV17 mitochondrial inner membrane protein	-0.23638073	Yes
UQCRB	ubiquinol-cytochrome c reductase binding protein	-0.22114263	Yes
NDUFS3	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	-0.211956	Yes
NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49kDa (NADH-coenzyme Q reductase)	-0.19109695	Yes
UQCRH	ubiquinol-cytochrome c reductase hinge protein	-0.17141747	Yes
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	-0.16190857	Yes
TIMM17A	translocase of inner mitochondrial membrane 17 homolog A (yeast)	-0.15715274	Yes
NDUFV1	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kDa	-0.13332075	Yes
ATP5G1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C1 (subunit 9)	-0.10680163	Yes
PHB	prohibitin	-0.07910502	Yes
NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	-0.058720794	Yes
RAB11FIP5	RAB11 family interacting protein 5 (class I)	-0.034326315	Yes
TOMM34	translocase of outer mitochondrial membrane 34	0.003430086	Yes

KEEG GLYCOLYSIS GLUCONEOGENESIS CASE X CONTROL t1			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
ACSS1	acyl-CoA synthetase short-chain family member 1	0.07301347	Yes

PGK1	phosphoglycerate kinase 1	0.14413437	Yes
PGAM2	phosphoglycerate mutase 2 (muscle)	0.21352457	Yes
PGM1	phosphoglucomutase 1	0.2759036	Yes
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	0.32415113	Yes
LDHA	lactate dehydrogenase A	0.3691173	Yes
GPI	glucose phosphate isomerase	0.4181434	Yes
ENO3	enolase 3 (beta, muscle)	0.4571662	Yes
BPGM	2,3-bisphosphoglycerate mutase	0.49810046	Yes
TPI1	triosephosphate isomerase 1	0.5033853	Yes
GALM	galactose mutarotase (aldose 1-epimerase)	0.5333864	Yes
PFKM	phosphofructokinase, muscle	0.5564322	Yes
PGAM4	phosphoglycerate mutase family member 4	0.5759253	Yes
FBP2	fructose-1,6-bisphosphatase 2	0.5952478	Yes
LDHB	lactate dehydrogenase B	0.582899	Yes
ENO1	enolase 1, (alpha)	0.602715	Yes
ALDOA	aldolase A, fructose-bisphosphate	0.62129945	Yes
ALDOC	aldolase C, fructose-bisphosphate	0.62718666	Yes
ACSS2	acyl-CoA synthetase short-chain family member 2	0.64731264	Yes
AKR1A1	aldo-keto reductase family 1, member A1 (aldehyde reductase)	0.59136206	No
PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	0.60139984	No
HK1	hexokinase 1	0.4881909	No
DLAT	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	0.4593283	No
ALDH3A2	aldehyde dehydrogenase 3 family, member A2	0.46381518	No
PDHA1	pyruvate dehydrogenase (lipoamide) alpha 1	0.47123164	No
PFKP	phosphofructokinase, platelet	0.43549097	No
PFKL	phosphofructokinase, liver	0.41168368	No
PGM2	phosphoglucomutase 2	0.4133238	No
ALDH7A1	aldehyde dehydrogenase 7 family, member A1	0.40297154	No
PDHB	pyruvate dehydrogenase (lipoamide) beta	0.3331222	No
PKLR	pyruvate kinase, liver and RBC	0.32172167	No
DLD	dihydrolipoamide dehydrogenase	0.28562146	No
HK3	hexokinase 3 (white cell)	0.23301776	No
ALDH9A1	aldehyde dehydrogenase 9 family, member A1	0.23804194	No
ALDH1A3	aldehyde dehydrogenase 1 family, member A3	0.1783741	No

REACTOME GLYCOLYSIS CASE X CONTROL t1

GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PGK1	phosphoglycerate kinase 1	0.10073286	Yes
PGAM2	phosphoglycerate mutase 2 (muscle)	0.2043479	Yes
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	0.27660838	Yes
GPI	glucose phosphate isomerase	0.34129986	Yes
ENO3	enolase 3 (beta, muscle)	0.40402585	Yes
PFKFB4	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	0.46630123	Yes
TPI1	triosephosphate isomerase 1	0.4921814	Yes
PFKM	phosphofructokinase, muscle	0.52397835	Yes
PFKFB1	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	0.50159955	Yes
ENO1	enolase 1, (alpha)	0.5306825	Yes
ALDOA	aldolase A, fructose-bisphosphate	0.5624456	Yes
ALDOC	aldolase C, fructose-bisphosphate	0.58073455	Yes
PFKFB2	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	0.49269098	No
PPP2CA	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	0.42801282	No
PPP2R1A	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), alpha isoform	0.41373107	No
PPP2R5D	protein phosphatase 2, regulatory subunit B (B56), delta isoform	0.41594967	No
PFKP	phosphofructokinase, platelet	0.36459467	No
PFKL	phosphofructokinase, liver	0.34349662	No
PKLR	pyruvate kinase, liver and RBC	0.24259447	No
PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), beta isoform	0.0180843	No
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	0.010247904	No

MOOTHA GLYCOGEN METABOLISM CASE X CONTROL t1			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
GYS1	glycogen synthase 1 (muscle)	0.076006	Yes
PPP1R3D	protein phosphatase 1, regulatory subunit 3D	0.035474524	Yes
PYGM	phosphorylase, glycogen; muscle (McArdle syndrome, glycogen storage disease type V)	0.11755047	Yes
AGL	amylo-1, 6-glucosidase, 4-alpha-glucanotransferase (glycogen debranching enzyme, glycogen storage disease type III)	0.18208763	Yes
GSK3B	glycogen synthase kinase 3 beta	0.25337955	Yes
SLC37A4	solute carrier family 37 (glycerol-6-phosphate transporter), member 4	0.32485268	Yes
PPP1CA	protein phosphatase 1, catalytic subunit, alpha isoform	0.37860835	Yes
PPP1R2	protein phosphatase 1, regulatory (inhibitor) subunit 2	0.42269337	Yes

PPP1R3A	protein phosphatase 1, regulatory (inhibitor) subunit 3A (glycogen and sarcoplasmic reticulum binding subunit, skeletal muscle)	0.4889237	Yes
PPP1CC	protein phosphatase 1, catalytic subunit, gamma isoform	0.36544198	No
GBE1	glucan (1,4-alpha-), branching enzyme 1 (glycogen branching enzyme, Andersen disease, glycogen storage disease type IV)	0.36072287	No
GAA	glucosidase, alpha; acid (Pompe disease, glycogen storage disease type II)	0.3116328	No
PPP1CB	protein phosphatase 1, catalytic subunit, beta isoform	0.2176989	No
GYG1	glycogenin 1	0.22291675	No
PYGL	phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI)	0.052709203	No
PHKA2	phosphorylase kinase, alpha 2 (liver)	0.013395293	No

MOOTHA GLYCOGEN METABOLISM CASE t1 X CASE t2			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PPP1R3D	protein phosphatase 1, regulatory subunit 3D	0.16643764	Yes
GYS1	glycogen synthase 1 (muscle)	0.27559268	Yes
PPP1CB	protein phosphatase 1, catalytic subunit, beta isoform	0.31016192	Yes
PPP1R3A	protein phosphatase 1, regulatory (inhibitor) subunit 3A (glycogen and sarcoplasmic reticulum binding subunit, skeletal muscle)	0.40646014	Yes
AGL	amylo-1, 6-glucosidase, 4-alpha-glucanotransferase (glycogen debranching enzyme, glycogen storage disease type III)	0.47425106	Yes
PYGM	phosphorylase, glycogen; muscle (McArdle syndrome, glycogen storage disease type V)	0.54834193	Yes
PPP1R2	protein phosphatase 1, regulatory (inhibitor) subunit 2	0.6143548	Yes
PPP1CA	protein phosphatase 1, catalytic subunit, alpha isoform	0.54543614	No
GAA	glucosidase, alpha; acid (Pompe disease, glycogen storage disease type II)	0.59190416	No
PHKA2	phosphorylase kinase, alpha 2 (liver)	0.25628397	No
PPP1CC	protein phosphatase 1, catalytic subunit, gamma isoform	0.19566712	No
GSK3B	glycogen synthase kinase 3 beta	0.19816622	No
SLC37A4	solute carrier family 37 (glycerol-6-phosphate transporter), member 4	0.19769964	No
GYG1	glycogenin 1	0.1696309	No
PYGL	phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI)	0.052225552	No

GBE1	glucan (1,4-alpha-), branching enzyme 1 (glycogen branching enzyme, Andersen disease, glycogen storage disease type IV)	0.03790926	No

REGULATION OF CYTOSKELETON ORGANIZATION AND BIOGENESIS CASE X CONTROL t1			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
ARHGEF2	rho/rac guanine nucleotide exchange factor (GEF) 2	0.03316236	Yes
NEXN	nexilin (F actin binding protein)	0.09743831	Yes
TSC1	tuberous sclerosis 1	0.1337783	Yes
MID1IP1	MID1 interacting protein 1 (gastrulation specific G12 homolog (zebrafish))	0.15742613	Yes
LATS1	LATS, large tumor suppressor, homolog 1 (Drosophila)	0.21806374	Yes
CDC42EP2	CDC42 effector protein (Rho GTPase binding) 2	0.26494676	Yes
ARHGEF10L	Rho guanine nucleotide exchange factor (GEF) 10-like	0.2945469	Yes
APC	adenomatosis polyposis coli	0.34008536	Yes
LIMA1	LIM domain and actin binding 1	0.38118127	Yes
KATNB1	katanin p80 (WD repeat containing) subunit B 1	0.3836476	Yes
GSN	gelsolin (amyloidosis, Finnish type)	0.40263125	Yes
CLASP1	cytoplasmic linker associated protein 1	0.41141662	Yes
MAPRE1	microtubule-associated protein, RP/EB family, member 1	0.4370639	Yes
CAPG	capping protein (actin filament), gelsolin-like	0.43520662	No
CLASP2	cytoplasmic linker associated protein 2	0.41874808	No
NF2	neurofibromin 2 (bilateral acoustic neuroma)	0.37957373	No
ARF6	ADP-ribosylation factor 6	0.36749858	No
RASA1	RAS p21 protein activator (GTPase activating protein) 1	0.36380446	No
NEBL	nebulette	0.37078935	No
MAPT	microtubule-associated protein tau	0.3084394	No
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)	0.3155228	No
CEP250	centrosomal protein 250kDa	0.29234973	No
NCK1	NCK adaptor protein 1	0.17506157	No
NCK2	NCK adaptor protein 2	0.07128077	No
SORBS3	sorbin and SH3 domain containing 3	0.07044592	No

NEGATIVE REGULATION OF CELLULAR COMPONENT ORGANIZATION AND BIOGENESIS CASE X CONTROL t1			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT

ARHGEF2	rho/rac guanine nucleotide exchange factor (GEF) 2	0.034818135	Yes
RTN4	reticulon 4	0.085308164	Yes
PACSIN3	protein kinase C and casein kinase substrate in neurons 3	0.15222615	Yes
TERF2IP	telomeric repeat binding factor 2, interacting protein	0.19961797	Yes
MID1IP1	MID1 interacting protein 1 (gastrulation specific G12 homolog (zebrafish))	0.23437274	Yes
RAC1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	0.27261388	Yes
APC	adenomatosis polyposis coli	0.2997492	Yes
LIMA1	LIM domain and actin binding 1	0.3417722	Yes
KATNB1	katanin p80 (WD repeat containing) subunit B 1	0.34504876	Yes
GSN	gelsolin (amyloidosis, Finnish type)	0.3647888	Yes
SET	SET translocation (myeloid leukemia-associated)	0.40582877	Yes
CLASP1	cytoplasmic linker associated protein 1	0.41731393	Yes
MAPRE1	microtubule-associated protein, RP/EB family, member 1	0.44361857	Yes
THY1	Thy-1 cell surface antigen	0.44530573	Yes
CAPG	capping protein (actin filament), gelsolin-like	0.47610074	Yes
TERF2	telomeric repeat binding factor 2	0.45130882	No
HMGB1	high-mobility group box 1	0.46613076	No
PAIP2	poly(A) binding protein interacting protein 2	0.43860796	No
ARF6	ADP-ribosylation factor 6	0.4339451	No
YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	0.41088012	No
ERCC1	excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)	0.40767527	No
CDC42	cell division cycle 42 (GTP binding protein, 25kDa)	0.34322494	No
ERCC4	excision repair cross-complementing rodent repair deficiency, complementation group 4	0.24674492	No
EIF2AK3	eukaryotic translation initiation factor 2-alpha kinase 3	0.10881366	No

ACTIN CYTOSKELETON ORGANIZATION AND BIOGENESIS **CASE X CONTROL t1**			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
PDPK1	3-phosphoinositide dependent protein kinase-1	0.033945695	Yes
AMOT	angiomotin	0.070539415	Yes
ABL1	v-abl Abelson murine leukemia viral oncogene homolog 1	0.09357531	Yes
NUAK2	NUAK family, SNF1-like kinase, 2	0.11775713	Yes
LIMK1	LIM domain kinase 1	0.13267873	Yes
DBN1	drebrin 1	0.14638913	Yes
ARHGEF2	rho/rac guanine nucleotide exchange factor (GEF) 2	0.16538118	Yes

MRAS	muscle RAS oncogene homolog	0.1667544	Yes
SORBS1	sorbin and SH3 domain containing 1	0.18894957	Yes
SHROOM1	shroom family member 1	0.2004331	Yes
ARPC4	actin related protein 2/3 complex, subunit 4, 20kDa	0.21009779	Yes
FGD1	FYVE, RhoGEF and PH domain containing 1 (faciogenital dysplasia)	0.22905678	Yes
TSC1	tuberous sclerosis 1	0.23837093	Yes
SSH1	slingshot homolog 1 (Drosophila)	0.2554579	Yes
MYH11	myosin, heavy chain 11, smooth muscle	0.2701921	Yes
DYNLL1	dynein, light chain, LC8-type 1	0.28441378	Yes
DSTN	destrin (actin depolymerizing factor)	0.29534206	Yes
FLNA	filamin A, alpha (actin binding protein 280)	0.3055571	Yes
KPTN	kaptin (actin binding protein)	0.3233233	Yes
LATS1	LATS, large tumor suppressor, homolog 1 (Drosophila)	0.32650152	Yes
CDC42EP2	CDC42 effector protein (Rho GTPase binding) 2	0.3297135	Yes
RAC1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	0.33929306	Yes
WASL	Wiskott-Aldrich syndrome-like	0.3385046	Yes
ARHGEF10L	Rho guanine nucleotide exchange factor (GEF) 10-like	0.35050374	Yes
RAC3	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	0.3553531	Yes
DLG1	discs, large homolog 1 (Drosophila)	0.3696003	Yes
MTSS1	metastasis suppressor 1	0.37901616	Yes
LIMA1	LIM domain and actin binding 1	0.38922855	Yes
CFL1	cofilin 1 (non-muscle)	0.39229783	Yes
ARPC1A	actin related protein 2/3 complex, subunit 1A, 41kDa	0.3871715	Yes
DST	dystonin	0.38670564	Yes
FGD2	FYVE, RhoGEF and PH domain containing 2	0.38633513	Yes
ROCK1	Rho-associated, coiled-coil containing protein kinase 1	0.39300364	Yes
GSN	gelsolin (amyloidosis, Finnish type)	0.39766422	Yes
RICTOR	-	0.40572113	Yes
MYOZ1	myozenin 1	0.40213743	No
NF1	neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)	0.37924445	No
CRK	v-crk sarcoma virus CT10 oncogene homolog (avian)	0.37131223	No
CAPG	capping protein (actin filament), gelsolin-like	0.37228256	No
WASF3	WAS protein family, member 3	0.37646213	No
MYH9	myosin, heavy chain 9, non-muscle	0.36927468	No
TAOK2	TAO kinase 2	0.37040472	No
ACTA1	actin, alpha 1, skeletal muscle	0.3579654	No

ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	0.36089736	No
SSH2	slingshot homolog 2 (Drosophila)	0.3512532	No
ARHGEF17	Rho guanine nucleotide exchange factor (GEF) 17	0.34981555	No
ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	0.33907065	No
NF2	neurofibromin 2 (bilateral acoustic neuroma)	0.33328462	No
WASF2	WAS protein family, member 2	0.33233544	No
RND3	Rho family GTPase 3	0.33210284	No
ARF6	ADP-ribosylation factor 6	0.3194002	No
FGD5	FYVE, RhoGEF and PH domain containing 5	0.30799395	No
BCAR1	breast cancer anti-estrogen resistance 1	0.312076	No
RASA1	RAS p21 protein activator (GTPase activating protein) 1	0.3131686	No
TTN	titin	0.30946496	No
NEBL	nebulette	0.31350935	No
LLGL1	lethal giant larvae homolog 1 (Drosophila)	0.30931765	No
WIPF1	WAS/WASL interacting protein family, member 1	0.26001358	No
CDC42BPG	CDC42 binding protein kinase gamma (DMPK-like)	0.25807148	No
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)	0.25365242	No
CDC42	cell division cycle 42 (GTP binding protein, 25kDa)	0.23018591	No
CDC42BPA	CDC42 binding protein kinase alpha (DMPK-like)	0.1892067	No
DOCK2	dedicator of cytokinesis 2	0.18374656	No
FLNB	filamin B, beta (actin binding protein 278)	0.17661224	No
CDC42BPB	CDC42 binding protein kinase beta (DMPK-like)	0.17109722	No
PREX1	-	0.16672099	No
ARPC5	actin related protein 2/3 complex, subunit 5, 16kDa	0.15549023	No
RHOJ	ras homolog gene family, member J	0.11759609	No
NCK1	NCK adaptor protein 1	0.107007734	No
FGD4	FYVE, RhoGEF and PH domain containing 4	0.09491257	No
PACSIN2	protein kinase C and casein kinase substrate in neurons 2	0.09496649	No
ARFIP2	ADP-ribosylation factor interacting protein 2 (arfaptin 2)	0.035752892	No
FSCN1	fascin homolog 1, actin-bundling protein (Strongylocentrotus purpuratus)	-0.002878757	No
RHOF	ras homolog gene family, member F (in filopodia)	0.005722532	No
EVL	Enah/Vasp-like	0.011573711	No
NCK2	NCK adaptor protein 2	0.018718315	No
TESK2	testis-specific kinase 2	0.004319156	No
TNXB	tenascin XB	0.009022218	No
SORBS3	sorbin and SH3 domain containing 3	0.00532494	No
ELMO1	engulfment and cell motility 1	0.01163458	No
RND1	Rho family GTPase 1	0.025040116	No

PRKG1	protein kinase, cGMP-dependent, type I	0.017584505	No
FGD6	FYVE, RhoGEF and PH domain containing 6	0.026243845	No

CYTOSKELETON ORGANIZATION AND BIOGENESIS **CASE X CONTROL t1**			
GENE	GENE_TITLE	RUNNING ES	CORE ENRICHMENT
TPPP	-	0.03288339	Yes
PDPK1	3-phosphoinositide dependent protein kinase-1	0.050044857	Yes
AMOT	angiomotin	0.070103094	Yes
ABL1	v-abl Abelson murine leukemia viral oncogene homolog 1	0.07896637	Yes
NUAK2	NUAK family, SNF1-like kinase, 2	0.0899258	Yes
LIMK1	LIM domain kinase 1	0.09289135	Yes
PAK1	p21/Cdc42/Rac1-activated kinase 1 (STE20 homolog, yeast)	0.10423227	Yes
DBN1	drebrin 1	0.11140865	Yes
TUBGCP2	tubulin, gamma complex associated protein 2	0.124831244	Yes
ARHGEF2	rho/rac guanine nucleotide exchange factor (GEF) 2	0.13473904	Yes
PCLO	piccolo (presynaptic cytomatrix protein)	0.13753198	Yes
NEXN	nexilin (F actin binding protein)	0.1420253	Yes
MRAS	muscle RAS oncogene homolog	0.1535937	Yes
SORBS1	sorbin and SH3 domain containing 1	0.16619688	Yes
MYL6B	myosin, light chain 6B, alkali, smooth muscle and non-muscle	0.17731306	Yes
SHROOM1	shroom family member 1	0.18157506	Yes
KIF5B	kinesin family member 5B	0.19229221	Yes
ARPC4	actin related protein 2/3 complex, subunit 4, 20kDa	0.19494826	Yes
FGD1	FYVE, RhoGEF and PH domain containing 1 (faciogenital dysplasia)	0.20519032	Yes
TSC1	tuberous sclerosis 1	0.20608422	Yes
SSH1	slingshot homolog 1 (Drosophila)	0.21490462	Yes
YKT6	YKT6 v-SNARE homolog (S. cerevisiae)	0.2250897	Yes
MYH11	myosin, heavy chain 11, smooth muscle	0.23283455	Yes
MYL6	myosin, light chain 6, alkali, smooth muscle and non-muscle	0.24193592	Yes
5-Sep	septin 5	0.25157663	Yes
DYNLL1	dynein, light chain, LC8-type 1	0.26125437	Yes
DSTN	destrin (actin depolymerizing factor)	0.26436415	Yes
FLNA	filamin A, alpha (actin binding protein 280)	0.2669862	Yes
KPTN	kaptin (actin binding protein)	0.27721795	Yes
PAFAH1B1	platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa	0.2815223	Yes

MID1IP1	MID1 interacting protein 1 (gastrulation specific G12 homolog (zebrafish))	0.2853397	Yes
LATS1	LATS, large tumor suppressor, homolog 1 (Drosophila)	0.29335856	Yes
HOOK3	hook homolog 3 (Drosophila)	0.30259475	Yes
CDC42EP2	CDC42 effector protein (Rho GTPase binding) 2	0.29950082	Yes
APOE	apolipoprotein E	0.30807284	Yes
PLD2	phospholipase D2	0.31297514	Yes
RAC1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	0.32121584	Yes
WASL	Wiskott-Aldrich syndrome-like	0.3138553	Yes
TMED10	transmembrane emp24-like trafficking protein 10 (yeast)	0.32225296	Yes
ARHGEF10L	Rho guanine nucleotide exchange factor (GEF) 10-like	0.32836238	Yes
APC	adenomatosis polyposis coli	0.32838488	Yes
RAC3	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	0.33555984	Yes
DLG1	discs, large homolog 1 (Drosophila)	0.34360656	Yes
ANK3	ankyrin 3, node of Ranvier (ankyrin G)	0.3497158	Yes
MTSS1	metastasis suppressor 1	0.355401	Yes
LIMA1	LIM domain and actin binding 1	0.3595685	Yes
DMD	dystrophin (muscular dystrophy, Duchenne and Becker types)	0.36146313	Yes
RHOT1	ras homolog gene family, member T1	0.36857027	Yes
CFL1	cofilin 1 (non-muscle)	0.37309808	Yes
ARPC1A	actin related protein 2/3 complex, subunit 1A, 41kDa	0.3623623	Yes
DST	dystonin	0.35654652	Yes
KATNB1	katanin p80 (WD repeat containing) subunit B 1	0.3626624	Yes
KIF1B	kinesin family member 1B	0.3697526	Yes
TUBGCP6	tubulin, gamma complex associated protein 6	0.37366083	Yes
FGD2	FYVE, RhoGEF and PH domain containing 2	0.3727357	Yes
ROCK1	Rho-associated, coiled-coil containing protein kinase 1	0.3743642	Yes
GSN	gelsolin (amyloidosis, Finnish type)	0.3740774	Yes
ABI2	abl interactor 2	0.37834227	Yes
RICTOR	-	0.38400975	Yes
MARK4	MAP/microtubule affinity-regulating kinase 4	0.38024947	Yes
MYOZ1	myozenin 1	0.3822342	Yes
CLASP1	cytoplasmic linker associated protein 1	0.37660682	Yes
SNAP29	synaptosomal-associated protein, 29kDa	0.37588665	Yes
MAPRE1	microtubule-associated protein, RP/EB family, member 1	0.37703222	Yes
NF1	neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)	0.37311253	Yes
TUBGCP5	tubulin, gamma complex associated protein 5	0.37027234	Yes

SOD1	superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult))	0.37296537	Yes
RCC1	regulator of chromosome condensation 1	0.37804592	Yes
CRK	v-crk sarcoma virus CT10 oncogene homolog (avian)	0.3778033	Yes
THY1	Thy-1 cell surface antigen	0.37771603	Yes
SMC3	structural maintenance of chromosomes 3	0.38079202	Yes
CAPG	capping protein (actin filament), gelsolin-like	0.38554215	Yes
WASF3	WAS protein family, member 3	0.38595834	Yes
MYH9	myosin, heavy chain 9, non-muscle	0.37516308	No
TAOK2	TAO kinase 2	0.37283552	No
SMC1A	structural maintenance of chromosomes 1A	0.3624592	No
CLASP2	cytoplasmic linker associated protein 2	0.36462402	No
ACTA1	actin, alpha 1, skeletal muscle	0.36610198	No
ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	0.3659102	No
ARHGAP4	Rho GTPase activating protein 4	0.369791	No
KIF13B	kinesin family member 13B	0.36446396	No
LASP1	LIM and SH3 protein 1	0.36284262	No
SSH2	slingshot homolog 2 (Drosophila)	0.36581233	No
ARHGEF17	Rho guanine nucleotide exchange factor (GEF) 17	0.3615478	No
SPTBN4	spectrin, beta, non-erythrocytic 4	0.36246285	No
PEX1	peroxisome biogenesis factor 1	0.36341113	No
ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	0.35575497	No
DES	desmin	0.34908554	No
NF2	neurofibromin 2 (bilateral acoustic neuroma)	0.35090283	No
WASF2	WAS protein family, member 2	0.34756207	No
RND3	Rho family GTPase 3	0.34501284	No
SNAP23	synaptosomal-associated protein, 23kDa	0.34425005	No
NUBP1	nucleotide binding protein 1 (MinD homolog, E. coli)	0.34702995	No
STMN1	stathmin 1/oncoprotein 18	0.33778834	No
ARF6	ADP-ribosylation factor 6	0.33933508	No
KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	0.33023727	No
FGD5	FYVE, RhoGEF and PH domain containing 5	0.3286789	No
BCAR1	breast cancer anti-estrogen resistance 1	0.33092526	No
RASA1	RAS p21 protein activator (GTPase activating protein) 1	0.3301985	No
RAN	RAN, member RAS oncogene family	0.3321233	No
TTN	titin	0.3272826	No
NEBL	nebulette	0.3296204	No
ARHGAP10	Rho GTPase activating protein 10	0.32907704	No
LLGL1	lethal giant larvae homolog 1 (Drosophila)	0.32614556	No
UXT	ubiquitously-expressed transcript	0.30591947	No

WIPF1	WAS/WASL interacting protein family, member 1	0.2774688	No
TNNT2	troponin T type 2 (cardiac)	0.27880004	No
CDC42BPG	CDC42 binding protein kinase gamma (DMPK-like)	0.27599272	No
MAPT	microtubule-associated protein tau	0.27104625	No
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)	0.27194208	No
MAP1S	microtubule-associated protein 1S	0.25153556	No
CDC42	cell division cycle 42 (GTP binding protein, 25kDa)	0.24864152	No
CEP250	centrosomal protein 250kDa	0.2464166	No
KIFAP3	kinesin-associated protein 3	0.2463104	No
KLHL20	kelch-like 20 (Drosophila)	0.23248412	No
MAP3K11	mitogen-activated protein kinase kinase kinase 11	0.23203112	No
NUSAP1	nucleolar and spindle associated protein 1	0.22319677	No
DYNC1I1	dynein, cytoplasmic 1, intermediate chain 1	0.2207803	No
MYO9B	myosin IXB	0.21821648	No
CKAP5	cytoskeleton associated protein 5	0.21597366	No
MYH10	myosin, heavy chain 10, non-muscle	0.21150638	No
CDC42BPA	CDC42 binding protein kinase alpha (DMPK-like)	0.21136378	No
DOCK2	dedicator of cytokinesis 2	0.2058683	No
FLNB	filamin B, beta (actin binding protein 278)	0.19857502	No
CDC42BPB	CDC42 binding protein kinase beta (DMPK-like)	0.19282143	No
PREX1	-	0.18814115	No
ARPC5	actin related protein 2/3 complex, subunit 5, 16kDa	0.17640477	No
CENPJ	centromere protein J	0.16994065	No
CNTROB	centrobin, centrosomal BRCA2 interacting protein	0.16744404	No
MYO1E	myosin IE	0.16675504	No
RHOJ	ras homolog gene family, member J	0.1398397	No
NCK1	NCK adaptor protein 1	0.1279464	No
KIF3B	kinesin family member 3B	0.12822892	No
FGD4	FYVE, RhoGEF and PH domain containing 4	0.116065934	No
PACSIN2	protein kinase C and casein kinase substrate in neurons 2	0.11458786	No
RHOT2	ras homolog gene family, member T2	0.10045373	No
MYO6	myosin VI	0.10237117	No
MARK1	MAP/microtubule affinity-regulating kinase 1	0.08314791	No
LRPPRC	leucine-rich PPR-motif containing	0.08071049	No
OPA1	optic atrophy 1 (autosomal dominant)	0.07919062	No
ARFIP2	ADP-ribosylation factor interacting protein 2 (arfaptin 2)	0.06683145	No
LRMP	lymphoid-restricted membrane protein	0.059105266	No
FSCN1	fascin homolog 1, actin-bundling protein (Strongylocentrotus purpuratus)	0.028137863	No

RHOF	ras homolog gene family, member F (in filopodia)	0.03299651	No
EVL	Enah/Vasp-like	0.035020374	No
NCK2	NCK adaptor protein 2	0.03828239	No
MYH6	myosin, heavy chain 6, cardiac muscle, alpha (cardiomyopathy, hypertrophic 1)	0.030611664	No
STX5	syntaxin 5	0.03410283	No
TESK2	testis-specific kinase 2	0.031171339	No
TNXB	tenascin XB	0.03098458	No
SORBS3	sorbin and SH3 domain containing 3	0.021640798	No
ELMO1	engulfment and cell motility 1	0.022043116	No
RND1	Rho family GTPase 1	0.029565593	No
PRKG1	protein kinase, cGMP-dependent, type I	0.014540377	No
FGD6	FYVE, RhoGEF and PH domain containing 6	0.014791662	No
MAP7	microtubule-associated protein 7	0.025527434	No

Chapter 4

Evaluation of a potential PSSM 2 locus on equine chromosome 18

Summary

Polysaccharide Storage Myopathy (PSSM) is a form of glycogen storage disease in horses, characterized by abnormal polysaccharide inclusions in skeletal muscle. One form of PSSM, termed PSSM1 is caused by a dominant mutation in the *GYS1* gene. A second form of PSSM, termed PSSM2, has also been described. The mutation responsible for PSSM2 is unknown. Histopathology and ultrastructural studies have demonstrated phenotypic differences between the *GYS1* (PSSM1) and non-*GYS1* (PSSM2) forms of PSSM.

Preliminary GWAS data revealed significant markers associated with PSSM2 on equine chromosome 18. The region contains 2 protein coding genes (*ACVR2A* and *ORC4L*) and a novel protein coding mRNA containing several repetitive elements (composed 100% of LINE sequences). Sanger sequence of *ORC4L* and most of *ACVR2A* revealed no mutations associated with PSSM2. However, due to poor assembly and annotation of the reference genome, the sequence of the 5' UTR/ promoter region and the first exon of *ACVR2a* could not be evaluated.

Next generation sequencing of amplicons generated by long PCR, next generation sequencing and *de novo* assembly of BAC clones, and next generation whole genome sequencing were performed in PSSM2 cases and controls to further interrogate this region on chromosome 18. The sequenced reads were mapped to the equine reference genome and the *de novo* reference created from the BAC clones. Variants were then called and annotated for location, biological function and abundance in cases versus controls. The use of BAC clones was efficient in creating a new reference encompassing the gaps present in the reference genome and the use of long PCR enabled good sequence comparison of the PSSM2 and controls. A total of 12,903,963 variants were called by GATK in cases and controls, in the entire genome. Mapping of the amplicons generated a total of 24,495 variants. In both cases the majority of the variants were intergenic. No variants significantly associated with PSSM2 were identified within this region on equine chromosome 18. The absence of variants that explain PSSM2 on chromosome 18 and the lack of other significant hits in the GWAS in other regions of the genome may be due to false positive association, low power of the GWAS analysis, partial penetrance of risk alleles on ECA18.

Introduction

Polysaccharide Storage Myopathy (PSSM) is a muscle glycogen storage disease in horses characterized by accumulation of excess and abnormal skeletal muscle glycogen and clinical myopathy[2]. A dominant mutation in the *GYS1* gene has been described as a causative mutation in one form of PSSM (PSSM1). However, a significant portion of PSSM horses cannot be attributed to the *GYS1* mutation and evidence points to a second heritable form of PSSM[25], referred to as type 2 PSSM. Histopathology and ultrastructural studies have demonstrated phenotypic differences between the *GYS1* (PSSM1) and non-*GYS1* (PSSM2) forms of PSSM[12,25].

Preliminary GWAS data performed in a cohort of 104 PSSM2 Quarter Horses (QH) cases and 124 QH controls, genotyped with the Illumina Equine SNP50 Beadchip, demonstrated six SNPs on equine chromosome 18 (ECA18) (p=9.44 x 10-7 to 2.84 x 10-6) and a single SNP on ECA3 (2.99 x 10-6) that were associated with PSSM2 using a dominant logistic regression model (Fritz *et al*, unpublished data, **figure 11**). The SNP on ECA3 was no longer associated with PSSM2 when correction for population structure was performed. Association of PSSM2 with the SNPs on chromosome 18 persisted after correction for multiple testing using 10,000 label-swapping permutations. These six ECA18 SNPs span an interval of ~350 kb (349,806 base pairs; chr18:30,479,171-30,828,977).

The 350 kb interval contains two protein coding genes, both expressed in equine skeletal muscle, activin A receptor, type IIA (*ACVR2A*) and origin recognition complex, subunit 4 like (*ORC4L*), as well as a novel transcript (ENSECAG00000024833) **(figure 11).** PCR was attempted to amplify 5' and 3' untranslated regions (UTR), coding (exonic) regions and exon/intron boundaries (splice sites) from each gene to enable Sanger sequencing and detection of simple mutations (nonsynonymous substitutions, missense mutations and small insertions or deletions). Complete sequencing of these regions in *ORCL4* did not identify simple mutations associated with PSSM2 (unpublished data). Sequencing of most *ACVR2A* exons and intron/exon boundaries has also not identified simple mutations within these sequences (unpublished data). Due to poor annotation and assembly of the reference equine genome (EqCab 2.0 [Twilight])[55], the

amplification and sequencing of the 5' UTR/ promoter region and the first exon of *ACVR2a* was unsuccessful. Attempts to design primers and "walk" through the region with Sanger sequencing of genomic DNA in PSSM2 cases and controls and BAC clones (bacterial artificial chromosome clones)[56] were also unsuccessful and suggest that large structural variants may be present in this region (unpublished data) (**figure 11**). The novel gene (ENSECAG00000024833) is approximately 68 kb in length, is 268kb from *ACVR2A* and is composed of several repetitive elements (LINE sequences).

The objective of this chapter was to use high throughput next-generation sequencing strategies to thoroughly interrogate the genomic regions on ECA18 associated with PSSM2 and overcome the issues with poor genome annotation and assembly in order to identify a potential genetic mutation that might be responsible for PSSM2. This approach allowed us: 1) to overcome limitations in SNP-based fine structure mapping and Sanger sequencing of candidate genes imposed by haplotype structure and incomplete genome annotation in the horse; and 2) to identify structural variants that may be difficult to detect with traditional Sanger sequencing technologies (i.e. large insertions, deletions, and/or inversions).

Material and Methods

To overcome poor annotation and possible structural variation associated with *ACVR2A*, first, 2 BAC clones that contain approximately 230 kb of equine sequence encompassing *ACVR2A* were sequenced to create a new assembly of the equine genome in this region. Second, long range PCR was used to amplify the entire 350 kb region associated with PSSM2 in 11 cases and 11 controls allowing for alignment and identification of all sequence differences (both simple and complex structural differences) between cases and controls across the entire region (**figure 11**).

PSSM2 cases were selected based on the presence of aggregates of amylase sensitive polyssacharide in the muscle biopsy, history of clinical signs of rhabdomyolysis

and the absence of the *GYS1* mutation. Controls were selected based on no prior history of rhabdomyolysis, normal muscle biopsy and absence of the *GYS1* mutation.

BAC clones next generation sequencing

Two BAC clones (CH241-325M2 - *BAC-A1* and CH241-57G18 - *BAC-A2*) encompassing the promoter region of *ACVR2A* gene were ordered from a BAC equine library originated from BRAVO, a male Thoroughbred closely related to Twilight (the mare that provided DNA for assembling of the current equine reference genome). The BAC clones were chosen based on their genomic location (chr18: 30,288,821 – 30,511,282 and chr18: 30,292,625 – 30,522,271), to cover the region around *ACVR2A* where the long-range PCR primers were unable to amplify (**figure 11; figure 12**). Those are regions where gaps are present in the reference genome. The BAC clones were sequenced using Illumina Mi-Seq sequencer, 150 bp paired-end run, 60x coverage.

Target next generation sequencing – equine chromosome 18

Long range PCR was used to amplify the entire 350 kb region associated with PSSM2, based on the GWAS, in 11 cases and 11 controls. 61 overlapping long-PCR primer pairs were designed and optimized to produce amplicons across a 537 kb (chr18: 30,251,565 – 30,789,048) region encompassing the 350 kb region around *ACVR2A* (**table**). The long range PCR was unable to amplify two small areas on *ACVR2A*, where gaps are present in the reference genome (**figure11**: dotted lines between primer pairs and **table 20:** primer pairs 23 and 27) and a third area around ENSECAG00000024833 (**figure11** and **table 20:** primer pair 52). The amplicons were sequenced at the University of Minnesota's Biomedical Genomics Center (BMGC) using the MiSeq technology. Each sample was prepared with an indexed barcode for a (150bp) paired–end run on the Illumina Mi-Seq sequencer (i.e. ~5 million reads). For the 537kb of amplified sequence, this resulted in > 60-fold sequence coverage per individual.

De novo assembly of BAC clones and target next generation sequencing reads

De novo assembly of the reads was performed on CLC bio[79], using de Brujin graphs. The software automatically finds the ideal k-mer size, after evaluating the amount

of input data. Prior to *de novo* assembly, Illumina adaptors were trimmed from the raw reads and reads with low quality were discarded. Next, the fastq reads were mapped to the BAC clone vector sequence (pTARBAC2.1) extracted from NCBI. Only unmapped reads (which would represent equine DNA) were used to create *de novo* contigs. The procedure above was first performed on each BAC clone individually (*BAC-A1and BAC-A2)* and then using both BAC clones reads together (*BAC-A1A2)*. The fasta *de novo* contigs were further assembled using the software Sequencher[80], allowing different overlapping and mismatch parameters. The best *de novo* contigs, based on length and blast identity to the region of interest on equine chromosome 18 were chosen for downstream analysis.

The same protocol was used for *de novo* assembly of targeted next generation sequencing reads of PSSM2 cases and controls, with the exception of mapping reads to the BAC clone sequence. Each one of the 11 PSSM2 cases and 11 controls (61 amplicons per horse) were assembled individually and the generated *de novo* contigs were further evaluated on Sequencher software to identify possible sequence differences (large structural variants) between cases and controls that be the causative mutation for PSSM2.

To identify possible variants surrounding the novel protein coding 'mRNA' (ENSECAG00000024833) and to overcome the same problems experienced when amplifying *ACVR2,* next generation whole genome sequence was performed in 3 PSSM2 horses and 3 controls.

Whole genome next generation sequencing

3 PSSM2 Quart Horse cases and 3 control Quarter Horses were selected for whole genome next generation sequencing (NGS). The cases were selected based on clinical signs of PSSM, severity of muscle biopsy findings and were also previously genotyped on the equine SNP50 array. Controls were randomly selected from the list of horses previously genotyped on the SNP50 array. Reads were sequenced using Illumina HiSeq, 100 bp paired-end, 12x coverage per horse (~160 million reads/horse). Each sample was sequenced in more than one lane.

Sequence Mapping and Variant Calling

Target next generation sequencing data on chromosome 18 and whole genome next generation sequencing were mapped and variants were called according to the protocol below.

Prior to mapping, the raw reads were checked for quality using FastQC, the ends were trimmed when the quality base score was below 20 and paired reads were synchronized. The Illumina fastq files were then aligned the equine reference genome (EqCab2) using BWA aligner, originating BAM files. Following mapping, variants were called by the Genome Analysis toolkit (GATK)[81,82]. Raw BAM files generated by BWA[83,84] are not of sufficient quality to identify real biological variants so quality filtering and correction for systematic biases was performed prior to variant calling. Reads that were ambiguously mapped (mapped equally well to multiple locations) and were of low quality were removed. Reads were then sorted in coordinate order with the reference genome, mates were synchronized and PCR duplicates were removed using Picard-Tools[85]. The GATK realigned target creator tool[81,86] was used to systematically check BAM files for the presence of indels and realign reads around them to avoid incorrect mapping and false positive SNPs, followed by base quality recalibration. Next, the GATK unified genotyper tool was used for variant discovery (SNPs and indels). Variants were then recalibrated by GATK's variant quality score recalibrator tool using only internal evidence (the data itself) and a final VCF file containing SNPs and indels was created. Each sample was mapped individually.

Cases and Controls were mapped individually using BWA and were then merged prior to GATK analysis, to call variants simultaneously. Sample identity was preserved during the process.

Variants from whole genome sequence were called initially with a confidence score threshold of Q20. The samples were individually sequenced at 12x coverage, but since depth of coverage is not uniformly around the entire genome, variants were also called using a quality score threshold of Q4, following guidelines by the Broad Institute

for shallow data (less than 10X coverage per sample). Variants from the target next generation sequencing were called with a confidence score threshold of Q20, since a coverage of more than 60X was expected per sample.

Miseq and whole genome NGS mapping to BAC clones

The reads originated from the target next generation sequencing on chromosome 18 and the whole genome next generation sequencing experiment were also mapped to the *de novo* assembly reference generated from the BAC clones using BWA.

Variant annotation and prioritization

After variants were identified from sequence data (target sequence on ECA18, and whole genome data), they were annotated by snpEff[87] software for variant location and their potential effect. SnPSift[88] case control was used to calculate variant abundance differences between cases and controls. Variant abundancy between groups was calculated for each one of the variants discovered by GATK. The program counts how many cases and controls have each one of those variants and calculates p-values for the differences in abundance using Fisher exact test. The p-values are calculated using different models: dominant, recessive, allelic and co-dominant, all of which are reported in the output. Variants were prioritized if 1) they were relatively abundant in cases compared to controls; 2) they occurred within a gene with known biologic function relevant to PSSM2 pathogenesis; and/or 3) the variant had a predicted functional effect. Variants detected on target reads and whole genome reads after mapping to the *de novo* reference created from BAC clones were evaluated for location and abundance between cases and controls.

Results

Target next generation sequencing

A total of 16,291,228 reads from the 61 amplicons were present in cases and controls; 99.98% of the reads mapped to EqCab2 and 15,110,584 (92.75% of total reads) mapped to equine chromosome 18. A total of 24,495 variants (SNPs and indels) were

called by GATK (3,176 on chromosome 18), with the vast majority being intergenic (**table 21**) single nucleotide polymorphisms (SNP) denoted as 'modifier' by GATK (usually non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of functional impact) (**table 22**). No variants with major effect or significant differences in abundance between the 11 cases and 11controls were observed in the region of interest on chromosome 18.

Whole genome next generation sequencing

A total of 168,074,580 reads were aligned to the reference genome (168,052,258 in pairs) and 12,903,963 variants (SNPs and indels) were identified across the entire genome (**table 23**). When evaluating only chromosome 18 (position: 30,200,171 - 31,027,348), 626 variants were called using a quality score of Q4 and 614 using a quality score of Q20. The vast majority of the variants were intergenic, single nucleotide polymorphisms and had a modifier effect (**table 24**). No variants with major effect or significant differences in abundance between cases and controls were observed in the region of interest on chromosome 18.

De novo assembly of BAC clones

De novo assembly of BAC clone CH241-325M2 (BAC-A1)

The input data for *de novo* assembly of BAC-A1 consisted of 505,153 fastq reads from next generation sequencing, from which 494,410 were matched when creating contigs, resulting in 616 contigs with N50 of 1,688 (minimum contig length: 210; maximum contig length: 157,687).

De novo assembly of BAC clone CH241-57G18 (BAC-A2)

The input data for *de novo* assembly of BAC-A2 consisted of 488,414 fastq reads from next generation sequencing, from which 480,080 were matched when creating contigs, resulting in 271 contigs with N50 of 75,820 (minimum contig length: 223; maximum contig length: 153,846).

De novo assembly of both BAC clones together (BAC-A1A2)

The input data for *de novo* assembly of BAC-A1A2 consisted of 993,567 fastq reads, from which 977,934 were matched when creating contigs, resulting in 1,295 contigs with N50 of 495 (minimum contig length: 204; maximum contig length: 157,687).

Further BAC clone assembly on Sequencher

The *de novo* contigs from CLCbio were further evaluated on Sequencher software in order to allow more flexibility with the overlapping parameters and the goal of creating even longer contigs. The longest contigs and the ones that blast to the equine reference chromosome 18 with higher identity were the ones selected from CLCbio for further evaluation on Sequencher. This process was done individually for contigs from *BAC-A1*, *BAC-A2* and *BAC-A1A2*. The 3 longest contigs from *BAC-A1* (sizes 157,689; 10,090 and 65,689) were further assembled on Sequencher, creating a final single contig of size 233,376 bp (**contig1**) (**figure 12**). Those 3 contigs overlapped by 26 and 16 base pairs, with only 2 mismatches (**figure 12**) and had 100% identity when blasted to equine chromosome 18 (chr18: 30,288,742 – 30,522,282). From *BAC-A2,* 2 contigs were further assembled together on sequencher (sizes 153,846 and 75,820) with a total size of 229,659. Those 2 contigs overlapped by just 7bp, with 1bp mismatch and had lower identity to equine chromosome 18. Contigs created from both BACs together (*BAC-A1A2*), had marked number of mismatches, less overlapping and slightly lower identity when assembled on Sequencher. Based on the amount of overlapping, best blast results to equine reference chromosome 18 and low number of mismatches, **contig1** was chosen as the *de novo* reference from the BAC clones for further mapping.

The *de novo* assembled sequence from **contig1** contained approximately 233kb through the promoter and 5' UTR of *ACVR2A* gene and encompasses 4 gap regions present on chromosome 18 in the reference genome. The gaps have a total size of 1087 bp (**figure 12**).

In addition to being mapped to the equine reference genome (**above**, EqCab2), both whole genome sequencing and targeted next generation sequencing products were 1) *de*

novo assembled to look for variants that were different between PSSM2 cases and controls; and 2) mapped to BAC ***contig 1***.

De novo assembly of target next generation sequencing reads

A total of 22 individual *de novo* assemblies were created (11 cases and 11 controls were assembled individually). Each individual sample had an average 816,544 read counts that were used to create the de novo contigs. After *de novo* assembly, the contigs had an average N50 of 6,430 (range: 2,378-18,844), the mean number of created contigs was 734 (range: 370-1,212) and their average minimum and maximum size was 161bp (range: 105-202) and 71,926bp (range: 27,169-134,684), respectively. All contigs were further evaluated on Sequencher software in attempt to identify possible structural variant differences between cases and controls. The expected size of each one of the contigs was 570kb (224kb and 208kb if accounting for amplification failure of 2 of the 61 primer pairs around *ACVR2A*; **figure 11**) and the longest individual de novo contig created by CLC was 130 kb, indicating that they did not contain the entire expected sequence. Cases and controls did not show any sequence differences (including structural variants) that could explain PSSM2, but this comparison was compromised by the fact that individual contigs, regardless of cases/control status were missing different sections of sequences in comparison to the reference.

*Target and whole genome NGS mapping to **contig1***

*Whole genome NGS mapping to **contig1***

The total number of input reads from the sequences of 3 PSSM2 and 3 control horses was 1,113,307,670, from where 2,906,220 mapped to **contig1,** resulting in the discovery of 438 variants.

*Target mapping to **contig1***

The total number of input reads from the targeted long PCR products (11 PSSM2 cases and 11 controls) was 388,924. 142,604 of these reads mapped to **contig1,** resulting in the discovery of 1,169 variants.

Seventy-five of the variants called by whole genome sequence mapping to contig1 were also called by target sequence (ie long PCR product) mapping to contig1. None of the variants called on the target and whole genome sequence data after mapping to contig1 had a significant difference in abundance between cases and controls, and could explain PSSM2. Overall, when comparing the number of variants that were called in the different experiments across the 233kb region in the reference genome (region that corresponds to contig1) and contig1 itself; the target sequencing mapping to contig1 produced the larger number of variants (1,168 variants), followed by target sequencing mapping to the reference genome (1,133 variants), whole genome sequencing mapping to the reference genome (677 variants) and whole genome sequencing mapping to contig1 (437 variants).

Discussion

The GATK unified genotyper did not identify any SNPs, or indels on equine chromosome 18 that clearly segregated with PSSM2. However, the unified genotyper is not ideal to call large indels. The largest indel called by the software in the present data was a 46bp deletion. To look for possible large structural variant differences between cases and controls, *de novo* assembly of the produced amplicons for each one of the cases and controls, individually, was performed. No large indels that could explain PSSM2 were observed between cases and controls, but the larger *de novo* contig (130kb) created by the *de novo* assembly software was smaller than its expected size of 570kb (based on the amplicons generated from long range PCR products; 224kb and 208kb if accounting for amplification failure of 2 of the primer pairs). This suggests that parts of sequences were missing, compromising any case-control comparisons for large structural variants. This issue can be explained by a poor performance of the *de novo* assembler software

and/or the difficulty with sequencing of the original amplicons (61 primer pairs). In both instances the presence of repetitive elements in the region was likely a limiting factor. The presence of gaps in the reference genome also made it difficult to successfully sequence a few of the long range PCR products. Also, no significant variant differences in cases and controls were observed when the amplicons were mapped to the equine reference genome, but problems with the original PCR products caused a few regions to have decreased coverage, again potentially compromising variant discovery. The 3 PSSM2 cases and 3 controls with whole genome sequence data provided the opportunity to detect variants that could have been missed by the target sequencing experiment, since the whole genome sequence data was not dependent on generation of specific PCR products to amplify this region of ECA18. Although large structural variants might not be called due to limitations of the variant caller software, evaluation of depth of coverage in the region of interest did not show evidence of the presence of large indels.

The *de novo* assembly generated from the BAC clones was successful in covering gaps in the reference genome and is a valuable approach to overcome issues with poor assembly of the equine reference genome. This process is facilitated by the presence of an equine BAC library. Mapping of the amplicons and whole genome sequence reads to the *de novo* contig created from the BAC clones also ruled out the possibility that important variants were not called in the region due to poor assembly or annotation of the reference genome.

No variants on chromosome 18 associated with PSSM2 were identified in the 22 horses analyzed. There are several possible technical explanations for this. One possible explanation is that incomplete penetrance or low-to-moderate relative risk for putative functional allele(s) on ECA18 made it difficult to differentiate risk alleles from other variants in the region. With only 22 horses, the allele count differences between cases and controls for low penetrance or low-to-moderate genotypic relative risk would not be easy to differentiate from random variants. Genotyping a much larger number of cases and controls for any putative functional variants would be needed to determine if low-to-moderate risk alleles for PSSM2 were present on ECA18. Failure to identify functional

alleles might also indicate that the initial association on ECA 18 was a false positive, and that the true region(s) of the genome harboring PSSM2 risk alleles were not identified due to low power of the GWAS analysis. The GWAS performed on the preliminary data had only 50,000 markers. While this technology was successful in finding several disease related markers in other studies, it was not powerful enough for PSSM2. Quarters Horses are known to have lower linkage disequilibrium (LD) in comparison with other breeds, due to a quick expansion of the breed. GWAS analysis relies on LD between the markers and the causative allele. A denser SNP genotyping array would be advantageous since it would increase the power to find a causative mutation. To overcome this issue in a cost-effective way, imputation can be performed and the density of the markers present on the initial GWAS can be increased without the need to further genotype the test population, as long as a reference population that was genotyped on a denser array is available.

It is also possible that other forms of excessive glycogen storage diseases, caused by more than 2 mutations are present, and the horses called PSSM2 actually represent more than one disease. Also, PSSM2 might not be caused by a single gene with major effect, but several genes with minor effect. In both cases the power of the GWAS would be significantly affected and an increased number of cases and controls would be necessary to find significant markers. If more than 2 forms of excessive glycogen storage disease are present in horses, a better understanding of the pathophysiology of those diseases would be necessary to better phenotype the horses prior to the analysis. Studying gene expression profile in PSSM1, PSSM2 horses and controls would help answering these questions.

In conclusion, no mutations that can explain PSSM2 were identified on equine reference chromosome 18. Further analysis, including the use of imputation, would be necessary in attempt to improve power of the initial GWAS analysis and identify variants responsible for the disease. The study of metabolic changes on equine muscle of PSSM2 horses prior and during exercise, and gene expression analysis would improve our understanding of the physiology of the disease and the differences between PSSM1 and PSSM2 at the molecular level.

Table 20. **Primer sequence of all 61 primer pairs.** F: forward; R: reverse; bp: base pair. Primers 23R, 27F and 27R did not work (due to gaps and repetitive elements in the reference genome). This corresponds to the region where the ordered BAC clones originated from. Primer pair 52F and 52R, located before the novel gene ENSECAG00000024833 also failed

Primer Name	Location	Forward	Reverse	Size	Gene Location
LR Set 1 F	30251565 30251586	GAAACCTCTACTTCCCCAGTCC			
LR Set 1 R	30260086 30260105		CGTTGTCACTTTCCCTCTGG	8540 bp	
LR Set 2 F	30259372 30259395	CTCTGCACCAATCAAGTACAGATG			
LR Set 2 R	30267774 30267797		CCACGCCATATGGTATAATGTCAG	8425 bp	
LR Set 3 F	30267128 30267148	ATCTTGCAGCTCAGGCTTACC			
LR Set 3 R	30276952 30276974		GTGCCAGGACAGAGATGAAATAC	9846 bp	
LR Set 4 F	30275904 30275926	GCTCGGTGCCTTTCACATAATAG			
LR Set 4 R	30285764 30285784		AATGAAAGCCTGGGTTGAGAG	9880 bp	
LR Set 5F	30284825 30284847	GCACATTTGGGGAGAAACTACAC			
LR Set 5R	30294550 30294570		ATGGCATCCCATTCTGTCAAG	9745 bp	
LR Set 6F	30293783 30293806	CATGTTACTTCTCACCCTCAGATG			
LR Set 6R	30302644 30302667		TCACCTATCCCTACTCTCTCAAGG	8884 bp	
LR Set 7F	30302246 30302266	GGAACCCAGAAAATGAAGCTC			
LR Set 7R	30311917 30311937		TACTTTGTTGGCCCCAATACC	9691 bp	
LR Set 8F	30311215 30311238	CACTCATGTAACCTCTGGGAGATG			
LR Set 8R	30320080 30320103		GAGAAAGGTAGTTGGAGAGTGGAC	8888 bp	
LR Set 9F	30319729 30319752	ATCACCCACAGGACTAACTGTTTC			
LR Set 9R	30328934 30328957		TTCTGCCCTACCCTCAACTATTTC	9228 bp	
LR Set 10F	30328238 30328258	CCCAACCTCTGGAATTACTGG			
LR Set 10R	30338048 30338068		GGATGGGGCAGAGGTATAATG	9830 bp	
LR Set 11F	30337195 30337217	AATTGGGCACAATATTCAGGATG			
LR Set 11R	30346536 30346559		TGTAAAACTCAGAGGTTGCAAGTG	9364 bp	
LR Set 12F	30345907 30345930	AGACAGACTTCACAGATTCAGTCC			
LR Set 12R	30355225 30355248		GTTTTCCCCCTATTCACTATCACC	9341 bp	
LR Set 13F	30354415 30354435	TCTTCCAGGAGACATGCAGAC			
LR Set 13R	30364137 30364158		AACCACTGTGGCCAAGTTAATC	9743 bp	
LR Set 14F	30363227-30363247	TGTCATGATGCCTGTCACCTC			
LR Set 14R	30372783-30372804		CTACTCTGAGGGTTTTGGCAAG	9577 bp	
LR Set 15F	30371974 30371994	AGGAGAGAAGGGTTGGGAAAC			
LR Set 15R	30381539-30381560		AGGGTAAAAAGTGTCCCTTTGC	9586 bp	
LR Set 16F	30379741 30379764	GTCCTTGTTAATATTGGCCATCAG			
LR Set 16R	30386275 30386298		CAGGTTCCATAATCCTCTGAGTTC	6557 bp	
LR Set 17F	30389665 30389685	ACAAAGGCGGTTTATCTGCTC			
LR Set 17R	30399071 30399092		TGAATTGACCTCATCCGTAATG	9427 bp	
LR Set 18F	30397929 30397949	CACAGATGCTGAACAGAGCAG			
LR Set 18R	30407802 30407822		TACCTACATGCCACACCCTTG	9893 bp	

LR Set 19F	30406820 30406840	TGGAGCCTAACCATCAAACAG			
LR Set 19R	30416481 30416501		TCTGGATCTCTGGTCATTTGC	9681 bp	
LR Set 20F	30415390 30415410	TCCGTGTTCTTCAGACACCAG			
LR Set 20R	30425352 30425372		ATGGATGCTGTGGAGGAGAAG	9982 bp	
LR Set 21F	30428928 30428950	GGAAAGCCTGATAATGAAAGGTC			
LR Set 21R	30433728 30433751		ACATCTTGTGTCTTTTTCCTCCTC	4823 bp	
LR Set 22F	30433156 30433176	AGAGCTCTGCAGAACACCATC			
LR Set 22R	30442872 30442892		ATCCCAAGGGTAACGTTTGAC	9696 bp	
LR Set 23Fa	30442537 30442560	GGACTGGTTTAGATAGCGAATCTC			ACVR2A Upstream
LR Set 23 Ra					ACVR2A Promoter
LR Set 23Fb	30446941 30446963	TTGATTTTACACCAGGAGGTTTG			ACVR2A Exon 1
LR Set 23Rb	30451338 30451361		TTGCATACTTCTTGTGGGAGTGAC	4420 bp	ACVR2A Intron 1
LR Set 24F	30450742 30450767	AAGGTATCAAGTTTAATTATATTTGC			ACVR2A Intron 1
LR Set 24R	30459957 30459983		TGTGAATCCTTCCTATTGTTCTCTACT	9241 bp	ACVR2A Intron 1
LR Set 25F	30459635 30459655	GGGCTGGTTTGGTAGAAAGAG			ACVR2A Intron 1
LR Set 25R	30469199 30469219		CTCGGCTCCTTTAATCCCTTC	9584 bp	ACVR2A Intron 1
LR Set 26F	30468981 30469003	TACACAAACCTAGCAGGGATAGC			ACVR2A Intron 1
LR Set 26 R	30476349 30476371		TTTACCTCAGACCCTACCTCTCC	7390 bp	ACVR2A Intron 1
LR Set 27F					ACVR2A Intron 1
LR Set 27R					ACVR2A Intron 1
LR Set 28F	30485921 30485944	TCTGTCTCTCCTTAATAGCCTTGC			ACVR2A Intron 1
LR Set 28R	30495682 30495707		CGGTTCTTACCTATAGTTCTCACCAC	9786 bp	ACVR2A Intron 2
LR Set 29F	30494729 30494749	TGTCTGCCAGATACCCAACTC			ACVR2A Intron 1
LR Set 29R	30504519 30504543		CCACCTTAACTCTAACCACAACAAG	9814 bp	ACVR2A Intron 4
LR Set 30FB	30504124 30504146	GAATTGCTGAATTGACTTCGTTG			ACVR2A Intron 4
LR Set 30RB	30514092 30514119		GCAGATAGGTACATCAAATACAAAGCAC	9995 bp	ACVR2A Intron 4
LR Set 31F	30512483 30512503	TGACTGCTGACACGTTCTTTG			ACVR2A Intron 4
LR Set 31R	30522107 30522127		AGCTGGGGAATCACAATCTAC	9644 bp	ACVR2A Intron 8
LR Set 32F	30521198 30521220	TCTTCTATCCTGTTGACCTCCAG			ACVR2A Intron 8
LR Set 32R	30530869 30530889		CAACAATTGTGAGGCACAGTC	9691 bp	End ACVR2A
LR Set 33F	30530020 30530039	TTGGTTTTGCTTCCACTTCG			ORC4L
LR Set 33R	30539843 30539864		ACCTCTCCAGTAACCGGGAAAC	9844 bp	ORC4L
LR Set 34F	30538816 30538837	CACCCAGAAACAACTCTCACTG			ORC4L
LR Set 34R	30548554 30548573		CGCAAAGCTTCTAATGTTGC	9757 bp	ORC4L
LR Set 35F	30547509 30547530	CCAGATATCCAATCATCCTCAC			ORC4L
LR Set 35R	30557408 30557428		CAGCCTGTTTTGCTAGACCAC	9919 bp	ORC4L
LR Set 36F	30556788 30556811	CCACTCAAGAACTTCATGTCTCAG			ORC4L
LR Set 36R	30564407 30564429		AGTGAATCGTTTGCTTTGCTTAG	7641 bp	ORC4L
LR Set 37F	30565110 30565134	CCCTGTTAGACTCATTGAGAGAGAG			ORC4L
LR Set 37R	30575021 30575044		ACTGTTTTCCGCAGCTACTATACC	9934 bp	ORC4L

LR Set 38F	30574244 30574265	GCTTCATGACATTTGACTTTGC			ORC4L
LR Set 38R	30584166 30584185		CGCAGCTGGATAAAAAGGAC	9941 bp	ORC4L
LR Set 39F	30583833 30583856	GACCTCCATGACAGGATAAATTGG			ORC4L
LR Set 39R	30592329 30592352		CTAGCTGGCTGAAAAACAGCTTAC	8519 bp	ORC4L
LR Set 40F	30591872 30591891	TTCATGGACGGAACTGAAAA			ORC4L
LR Set 40R	30601649 30601668		CTTGGATCGTGCTTTTGATG	9796 bp	ORC4L
LR Set 41F	30600765 30600784	TGTTCACAGATGGCACGATT			ORC4L
LR Set 41R	30610372 30610391		GCCTCGGATTTAGAGCCTTT	9626 bp	ORC4L
LR Set 42F	30609561 30609580	GGAATGCTGACAGAGGTGGT			ORC4L
LR Set 42R	30619125 30619144		GCCAAGAAATCCTTCAGCAC	9583 bp	ORC4L
LR Set 43F	30618196 30618215	AGGTGAGAACCAGGCTTGAC			
LR Set 43R	30627848 30627867		TCCCCTTGAGTCAGGATTTG	9671 bp	
LR Set 44F	30626728 30626749	TGAACTTTCGAAAAGCACTAGG			
LR Set 44R	30636543 30636564		TTTGGGAACAACTAACTTCAGC	9836 bp	
LR Set 45F	30635736 30635755	CACCATAGTGCTCCGGACTG			
LR Set 45R	30645328 30645347		TCTTGTGGGCTGTCTTAGGG	9611 bp	
LR Set 46F	30643793 30643813	TGGATTTTGTATCCCCTTTCC			
LR Set 46R	30654237 30654257		TGCCATCACAGGGTTAATCAC	10464 bp	
LR Set 47F	30653511 30653532	GGGGCTAGAATCTCTAGTGCTG			
LR Set 47R	30662944 30662963		GATGGCAAGGAATCTGCAAC	9452 bp	
LR Set 48F	30662137 30662158	TGGTGGTCTTATCCTTGAGATG			
LR Set 48R	30671683 30671702		TTGGGAGGATTCTTGGTCAG	9560 bp	
LR Set 49F	30670678 30670701	AAGTCTACCCTGCTCATGCTATCC			
LR Set 49R	30680598 30680621		CTATCACCTCTCTGCGCTACTTAC	9550 bp	
LR Set 50F	30679787 30679807	GCTGGAAATGTATCAAAATGC			
LR Set 50R	30689373 30689392		ATGCCAATGCACTTCTTTGA	9605 bp	
LR Set 51F	30688570 30688593	ATTCTCCTTATGGCACCTAATCTG			
LR Set 51R	30694509 30694531		TGGTTACTTCTCAGCAGGTTTTG	5961 bp	
LR Set 52F					
LR Set 52R					
LR Set 53F	30706564 30706588	CCTTTGACTTTGGTATTCCATTAGG			
LR Set 53R	30715505 30715528		TCTTTCCCATCTAGGTCTGTTTTC	8964 bp	
LR Set 54F	30715056 30715080	AAGTAATTACCCACAAGAGAAGCT G			
LR Set 54R	30724139 30724163		CTTTCTTTGAGTAAGGAATGTGTGG	9107 bp	
LR Set 55F	30723804 30723827	CCTCCATATGTGATAACCAGAGTG			
LR Set 55R	30732857 30732880		GGGATATGCTGAGACTCTAGAAGG	9076 bp	
LR Set 56F	30732064 30732087	CAGAACCTCCATCTGTTTTCTATG			
LR Set 56R	30741773 30741796		TCTGAAGAAATGACGACTGAAAGC	9732 bp	
LR Set 57F	30741123 30741146	TATTTCTTGGGCAAAACTCCTAGC			
LR Set 57R	30750366 30750389		CTGCAGTAGTTTTTATGGTGGATG	9266 bp	
LR Set 58F	30749675 30749698	TATCCTTGGTCCCTAACATAGTGC			
LR Set 58R	30758913 30758936		AAAGGCTAGGATTCAGGTTTCTTC	9261 bp	

LR Set 59F	30757957 30757980	CTGCAAAATGGGAACCTAAAACAC		
LR Set 59R	30768916 30768942		GTTCTCATAGACCACCTTAATTACATC	10985 bp
LR Set 60F	30778657 30778679	CTTTCGTCTTTCCAATACCTGTG		
LR Set 60R	30785283 30785305		ACAATTGTTGAGACAGGAAGGTC	6648 bp
LR Set 61F	30778522 30778545	TGATCTTGGTTCTAACAACAGTGC		
LR Set 61R	30789025 30789048		AGACAGAGCCAAGCAGAACAGTAG	10526 bp

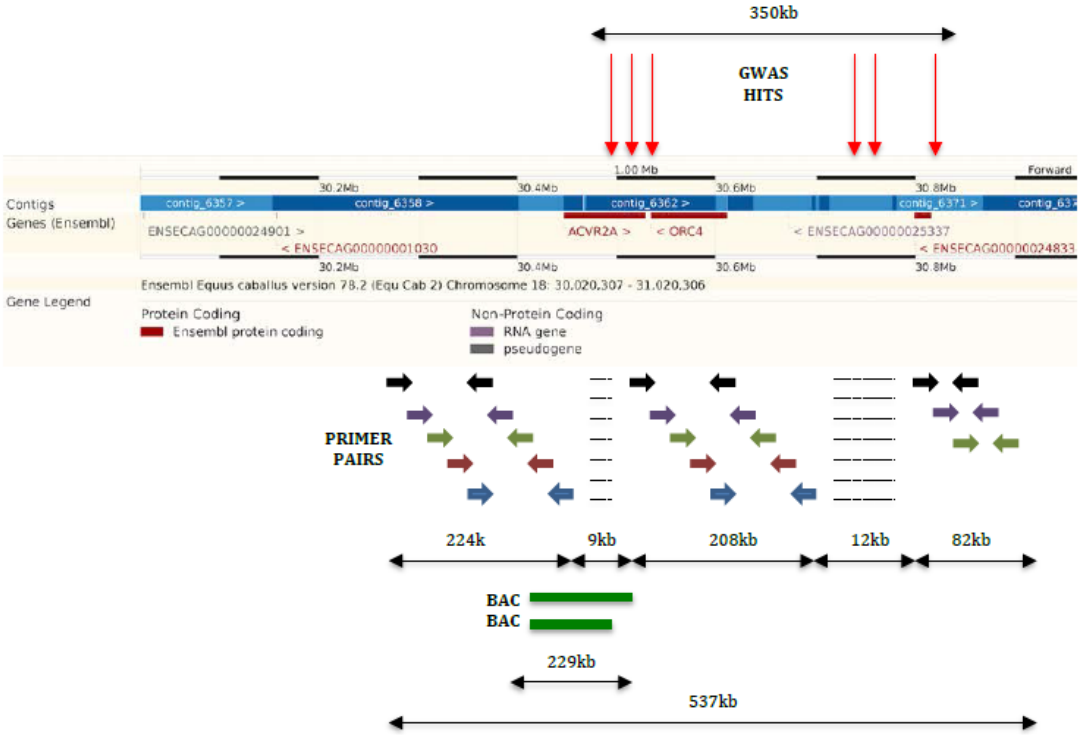


Figure 11. Ensembl. Equine chromosome 18: the vertical red arrows represent the 6 significant SNPs on the GWAS analysis; horizontal colorful arrows represent the approximate genomic position of the 61 PCR products in relation to the genome; green boxes represent the genomic location of the BAC clones in relation to the genome; horizontal dotted lines represent regions that could not be sequenced with the long range PCR; black horizontal arrows represent the length of the PCR products, BAC clones and GWAS hits.

Number of effects by type and region

Type

Type (alphabetical order)	Count	Percent
CODON_CHANGE_PLUS_CODON_DELETION	6	0.021%
DOWNSTREAM	1,338	4.585%
EXON	42	0.144%
FRAME_SHIFT	31	0.106%
INTERGENIC	14,177	48.578%
INTRON	8,945	30.65%
NONE	2,730	9.354%
NON_SYNONYMOUS_CODING	312	1.069%
SPLICE_SITE_ACCEPTOR	7	0.024%
SPLICE_SITE_DONOR	12	0.041%
START_GAINED	1	0.003%
START_LOST	1	0.003%
STOP_GAINED	9	0.031%
SYNONYMOUS_CODING	153	0.524%
UPSTREAM	1,377	4.718%
UTR_3_PRIME	38	0.13%
UTR_5_PRIME	5	0.017%

Region

Type (alphabetical order)	Count	Percent
DOWNSTREAM	1,338	4.585%
EXON	554	1.898%
INTERGENIC	14,177	48.578%
INTRON	8,945	30.65%
NONE	2,730	9.354%
SPLICE_SITE_ACCEPTOR	7	0.024%
SPLICE_SITE_DONOR	12	0.041%
UPSTREAM	1,377	4.718%
UTR_3_PRIME	38	0.13%
UTR_5_PRIME	6	0.021%

Table 21. **Target next generation sequencing – distribution of variants called by GATK.** The vast majority of the variants were intergenic

Number of changes by type

Type	Total
SNP	22,595
INS	769
DEL	1,131
Total	24,495

Number of effects by functional class

Type	Total
MISSENSE	313 (65.89%)
NONSENSE	9 (1.89%)
SILENT	153 (32.21%)

Number of effects by impact

Type	Total
HIGH	60
LOW	154
MODERATE	318
MOFIFIER	28,652

Table 22. **Target next generation sequencing – distribution of variants called by GATK** by type, effect by functional class and effects by impact.

Number of effects by type and region

Type

Type (alphabetical order)	Count	Percent
CODON_CHANGE_PLUS_CODON_DELETION	90	0.001%
CODON_CHANGE_PLUS_CODON_INSERTION	63	0%
CODON_DELETION	124	0.001%
CODON_INSERTION	106	0.001%
DOWNSTREAM	705,109	4.616%
EXON	16,707	0.109%
EXON_DELETED	2	0%
FRAME_SHIFT	6,655	0.044%
INTERGENIC	8,762,874	57.372%
INTRON	3,982,642	26.075%
NONE	914,400	5.987%
NON_SYNONYMOUS_CODING	46,474	0.304%
NON_SYNONYMOUS_START	13	0%
SPLICE_SITE_ACCEPTOR	1,747	0.011%
SPLICE_SITE_DONOR	1,941	0.013%
START_GAINED	612	0.004%
START_LOST	56	0%
STOP_GAINED	533	0.003%
STOP_LOST	33	0%
SYNONYMOUS_CODING	50,406	0.33%
SYNONYMOUS_START	1	0%
SYNONYMOUS_STOP	26	0%
UPSTREAM	769,798	5.04%
UTR_3_PRIME	7,810	0.051%
UTR_5_PRIME	5,576	0.037%

Region

Type (alphabetical order)	Count	Percent
DOWNSTREAM	705,109	4.616%
EXON	121,289	0.794%
INTERGENIC	8,762,874	57.372%
INTRON	3,982,642	26.075%
NONE	914,400	5.987%
SPLICE_SITE_ACCEPTOR	1,747	0.011%
SPLICE_SITE_DONOR	1,941	0.013%
UPSTREAM	769,798	5.04%
UTR_3_PRIME	7,810	0.051%
UTR_5_PRIME	6,188	0.041%

Table 23. **Whole genome next generation sequencing – distribution of variants called by GATK in the entire genome.** The vast majority of the variants were intergenic

Number of changes by type

Type	Total
SNP	11,622,781
INS	602,815
DEL	678,316
Total	12,903,963

Number of effects by functional class

Type	Total
MISSENSE	46,574 (47.75%)
NONSENSE	522 (0.53%)
SILENT	50,434 (51.71%)

Number of effects by impact

Type	Total
HIGH	10,967
LOW	51,058
MODERATE	46,857
MOFIFIER	15,164,916

Table 24. **Whole genome next generation sequencing – distribution of variants called by GATK** by type in the entire genome, effect by functional class and effects by impact

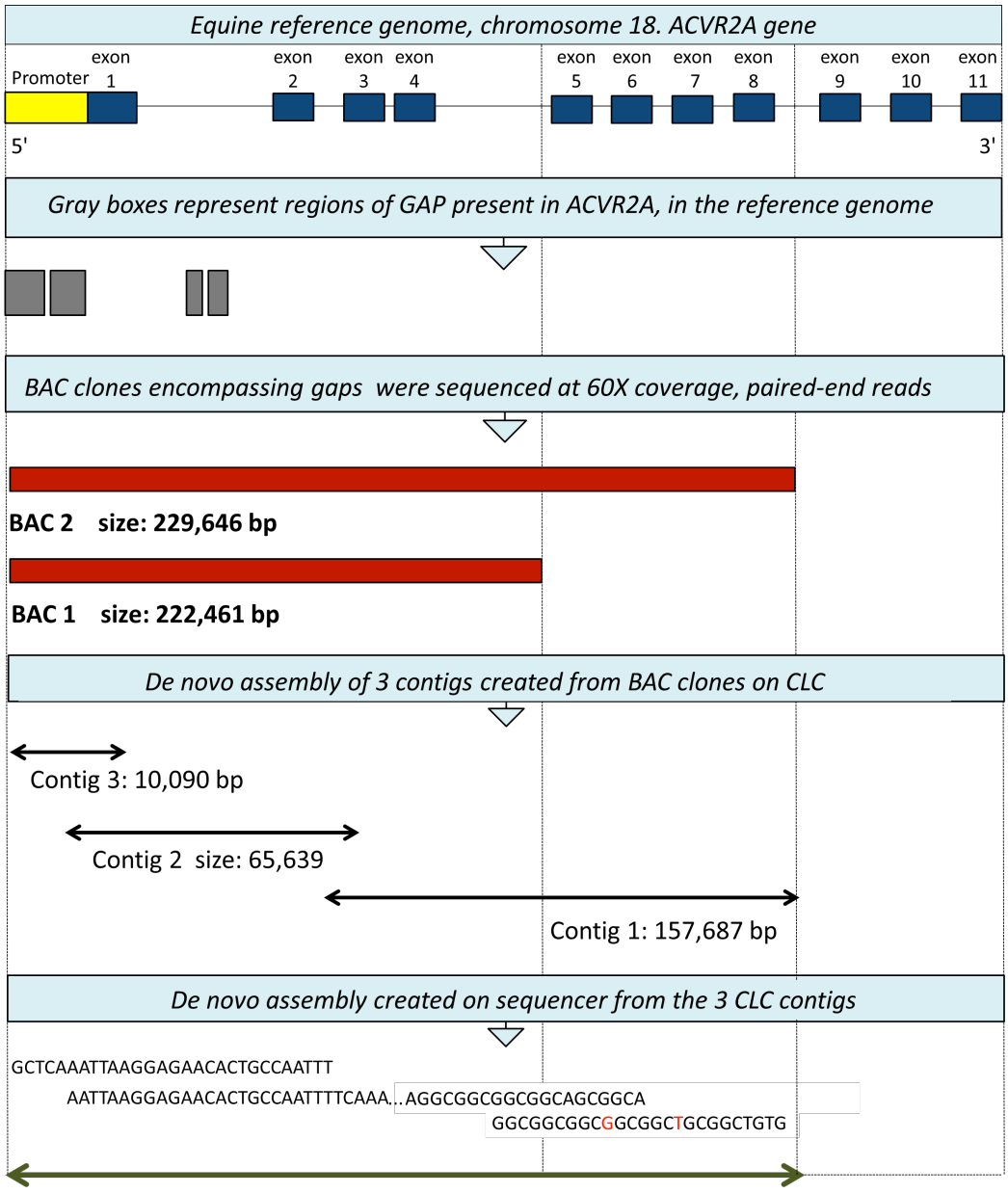


Figure 12. Flow chart of the *de novo* assembly of BAC clones. The gene structure of the reference *ACVR2A* and the gaps (gray boxes) present in the reference genome are represented, followed by the BAC clones that were ordered (red boxes) and their location in relation to *ACVR2A*. The 3 longest contigs used to generate contig1 from BAC-A1 on CLCbio are represented by the black arrows. Those 3 contigs were further assembled on Sequencher, creating the new reference for the region (contig1 – green arrow),that encompasses all the gaps present in the reference genome in the region.

Chapter 5

Enhancing the PSSM2 GWAS by Genotype Imputation

Summary

Polysaccharide Storage Myopathy (PSSM) is a glycogen storage disease in horses characterized by abnormal polysaccharide inclusions in skeletal muscle. There are two forms of PSSM in horses recognized so far, type 1 and type 2. A mutation in the *GYS1* gene results in type 1 PSSM, but the mutation responsible for type 2 PSSM (PSSM2) is unknown. Although preliminary genome wide association (GWA) data revealed significant markers associated with PSSM2 on equine chromosome 18 no variants associated with PSSM2 were identified after extensive investigation of this region using a combination of target and whole genome sequencing in PSSM2 cases and controls (Chapter 4). Failure to identify putative functional variants on ECA18 may be due to the limitations of the approach used to investigate the region, or due to incomplete penetrance of risk the allele(s) on ECA18; making it hard to differentiate from other variants in the region with only 22 horses. Failure to identify functional alleles might also indicate that the initial association on ECA 18 was a false positive, and that the true region(s) of the genome harboring PSSM2 risk alleles were not identified due to low power of the GWA analysis.

In this study, imputation was performed to increase the number of SNP markers in the initial GWA from 54,000 to close to 1.8 million markers (970,131 after pruning for minor allele frequency and marker missingness). Association using the imputed dataset, while accounting for population stratification, revealed new regions associated with PSSM2 on chromosomes 27 and 11. Subsequent Bayesian analysis of both single marker and haplotype associations using hapQTL supported the association only on chromosome 11. The region of ECA11 encompasses several annotated genes. Next generation sequencing data from cases and controls revealed non-synonymous mutations in the phosphoribosylformylglycinamidine synthase gene (*PFAS*) in 2 out of 3 cases and none of the controls. The *PFAS* gene encodes an enzyme that catalyzes the forth step of inosine monophosphate biosynthesis, which is part of the purine *de novo* nucleotide cycle, an important step for energy metabolism. Sequencing the *PFAS* gene in a larger population of cases and controls will be necessary to confirm an association with PSSM2. This

study therefore has identified a new region and positional candidate gene in which to focus our efforts to define the genetic basis for PSSM2.

Introduction

The power of many GWA studies in non-human species is limited by the density of the SNP markers on the commercially available genotyping arrays. The goal of genotype imputation is to predict the genotypes of SNPs that are not directly genotyped in the study sample. In imputation, a reference panel of haplotypes at a denser set of SNPs is used to impute the genotypes in a study sample of individuals that have been genotyped for only a subset of these SNPs. This is possible by using patterns of haplotypic variation seen in the reference panel[89]. The use of genotype imputation to increase marker density can be carried out across the whole genome to improve the power of GWA studies[58].

According to McCoy et al[59] the size of the imputed population did not impact imputation success when assessing genotype imputation accuracy from the 54,000 (SNP50) to 65,000 (SNP60) equine genotyping array. However, imputation accuracy did increase with larger reference population sizes and when imputed and reference populations were breed-matched. Linkage disequilibrium (LD) also influenced genotype accuracy, as breeds with longer LD (Thoroughbred) had higher imputation success than did breeds with shorter LD (Quarter Horses)[59]. Previous studies have also reported increased imputation accuracy with increased numbers of samples in the reference population, increased number of markers in the test set, and the presence of closely-related animals in the reference panel[90].

The previous chapter reported results of a preliminary GWA data performed on PSSM2 cases and controls, genotyped with low density (54,000 marker[57]) equine SNP array, that revealed significant markers associated with PSSM2 on equine chromosome 18 using a dominant logistic regression ($p=9.44 \times 10^{-7}$ to 2.84×10^{-6}). Sanger and next generation sequencing approaches and *de novo* assembly of BAC clones were performed to further interrogate the region on chromosome 18 in cases and controls, but no variants associated with PSSM2 were identified. Failure to identify putative functional variants on ECA18 may be due to the limitations of the approach used to investigate the region, or

due to incomplete penetrance of risk the allele(s) on ECA18; making it hard to differentiate from other variants in the region with only 22 horses. Failure to identify functional alleles might also indicate that the initial association on ECA 18 was a false positive, and that the true region(s) of the genome harboring PSSM2 risk alleles were not identified due to low power of the GWA analysis.

A new high density equine SNP array containing 670,000 markers is now available. The 54,000 markers from the equine SNP50[57] are also present in the new 670k equine SNP array. Haplotype information from 1.8 million markers (from which the 670K SNPs were selected from) is also available, making it possible to impute from 50k to 670K or even up to 1.8 million markers. Population based imputation methods rely on linkage disequilibrium relationship between SNPs, and essentially consist of two steps (inference of haplotypes and imputing untyped genotypes in the test set using information from the best fit haplotypes derived from the reference panel)[91]. Imputing less-dense genotype arrays is a cost-effective and powerful way to include more genetic markers into association studies without the need to re-genotype the samples[58]. Imputation from 50K to 800K has been previously reported with very high accuracy in dairy cattle[90].

The purpose of the present study is to impute from 54,000 to 1.8 million markers, perform association analyses of the imputed data, and look for possible causative mutations in any regions of GWA significance in the next generation sequencing data from 3 PSSM2 cases and 3 controls.

Material and Methods

Imputation from low to high-density markers

Description of the datasets

Two datasets genotyped on 2 different SNP chips were used for the imputation, the reference and the test population datasets. The reference population consisted of 63 Quarter Horses (QH) that were genotyped on the high-density 1.8 million Affymetrix

SNP array. The test population, to be imputed, consisted of 104 PSSM2 QH cases and 124 QH controls, previously genotyped on the lower density Illumina Equine SNP50 Beadchip (54,000 markers). Twenty-two of the horses genotyped on the 50K SNPchip were also genotyped on the 1.8 million array and are therefore present in the test and reference population, at different genotyping density. Also, nearly all of the 54,000 thousand markers from the SNP50 are present in the high-density array.

Data pre-processing and imputation

The ped and map files from the 50K chip test population were converted to variant calling format (VCF). The reference population VCF file was split by chromosome using vcftools software[92]. Each chromosome was then individually phased by BEAGLE4 software[93]. Prior to imputation the reference and test population VCF files were re-organized so they had the same chromosome and marker order. The test population was then imputed using data from the phased reference population by BEAGLE4 software[94]. Each chromosome was imputed individually.

BEAGLE is a population-based imputation technique that relies on LD relationships between markers. First there is inference of haplotypes followed by imputation of un-genotyped markers in the test population using information from the best-fit haplotype derived from the reference panel. The output file provides the probability of all three possible genotypes (homozygous wild-type, heterozygous and homozygous mutant) at each imputed marker. Further analysis was conducted using the most likely genotype and all 3 genotype probabilities (posterior probabilities)[95].

Imputation accuracy

To determine imputation accuracy and identify poorly imputed markers, the 22 horses that were genotyped on both arrays were extracted from both the imputed VCF file and the reference VCF file to generate 2 datasets; one with the imputed data from the 22 horses and one with their known genotype. Next the imputed variants were checked for accuracy, by comparing which genotype BEAGLE predicted at the specific location and the real genotype. SNPs that did not perform well (did not have the corrected genotype in

more than 90% of the horses) were removed from the VCF file used for downstream analysis (association test).

Genome wide association analysis (GWA)

The calculated genome inflation factor for this population of cases and controls was 1.12. Inflation factors greater than 1 are indicative of population structure. Also, multidimensional scaling plots to visualize the relationship between cases and controls in 6 different dimensions showed controls clustering separated from cases on PC3 (principal component 3) and PC4 and PC1 and PC3; another evidence for the presence of population stratification (figure 13; figure 14). For this reason an association analysis that takes into consideration population structure was performed.

Imputed file with all genotype probabilities and accounting for population structure

The Genome-wide Efficient Mixed Model Association algorithm software (GEMMA) was used for association analysis. It fits a univariate linear mixed model (LMM) for marker association tests with a single phenotype to account for population stratification[96]. First the imputed VCF file (containing only the markers remaining after filtering for accuracy) was converted to a genotype probability format file using vcf2probs from BEAGLE utilities. The file was then converted to bimbam format as require by GEMMA software[97]. By default, prior to running the association, the file was filtered for minor allele frequency (alleles with maf > 1% were kept in the analysis) and marker missingness (SNPs with a 95% genotyping rate were kept in the analysis). A relatedness matrix file in addition to both genotype and phenotype files, was created and contained the estimated relatedness between individuals. An association was then performed taking into consideration all 3 possible genotypes at each imputed marker and their probabilities and the calculated relatedness between individuals.

Haplotype analysis

A haplotype association test was performed with haplotype quantitative loci (hapQTL) software[98]. This program has advantages in comparison with other methods. First, it works directly with diploid data and no phasing is required, removing the issue of phasing uncertainty on downstream analysis. Secondly, most programs arbitrarily use predetermined window lengths to define haplotypes and then test each haplotype for association. HapQTL defines haplotypes by learning from the data using linkage disequilibrium (LD) information, which is a more appropriate way since the size of LD blocks varies along the genome. Each SNP is a core marker for its local haplotypes and its associated test statistics are computed[99].

The method first infers ancestral haplotypes and their loadings at each marker for each individual. The loadings are then used to quantify local haplotype sharing (LHS) between individuals at each marker. A statistical model links the local haplotype sharing and phenotypes to test for association. LHS reflects genetic similarity between individuals. By testing whether the genetic similarity is associated with the phenotypes, the program can identify associations between local haplotypes and phenotypes[100].

A window of 2 megabases (Mb) around the most significant markers found on the association analysis performed on GEMMA[101] was included in the haplotype analysis. The imputed ped and map files from cases and controls were converted to bimbam format using PLINK (--recode-bimbam)[102]. The analysis on hapQTL was performed as single marker and haplotype significance. Markers with minor allele frequency lower than 10% were removed from prior to the analysis. The program generates Bayes factors (the change of odds ratio in light of data) for the haplotype and single marker association analysis.

Variant discovery from whole genome next generation sequencing

A VCF file containing variant data from 3 PSSM2 cases and 3 controls from whole genome next generation sequencing (NGS) was available from the work performed around chromosome 18. This file was originally used to investigate possible variants relevant to PSSM2 according to the original GWAS data. Briefly, the cases were selected based on severity of muscle biopsy findings and were also previously genotyped on the equine SNP50 array. Controls were randomly selected from the list of horses

previously genotyped on the SNP50 array. Reads were sequenced using Illumina HiSeq, 100 bp paired-end, 12x coverage per horse (~160 million reads/horse). Each sample was sequenced in more than one lane. The reads were mapped to the equine reference genome (EqCab2) using BWA[103] aligner and variants were called by the Genome Analysis toolkit (GATK)[81] using the Unified genotyper tool for variant discovery (SNPs and indels). Cases and controls were mapped individually and then merged prior to GATK analysis, to call variants simultaneously. Variants were then annotated by snpEff software for variant location and their potential effect. SnPSift case control was used to calculate variant abundance differences between cases and controls. Variants were prioritized if 1) they were relatively abundant in cases compared to controls; 2) they occurred within a gene with known biologic function relevant to PSSM2 pathogenesis; and/or 3) the variant had a predicted functional effect.

Results

Imputation results and accuracy

The reference VCF file had approximately 1,846,962 markers. The total number of markers in the target population (PSSM2 cases and controls) after imputation was approximately 1,662,356. Markers that were not correctly imputed in at least 90% of the 22 horses present on target and reference population were removed from the final imputed file prior to downstream analysis. After removing markers for imputation accuracy, 1,295,518 were kept for downstream analysis; 84% of the markers analyzed had their genotype correctly imputed by BEAGLE

Association test, all genotype probabilities and accounting for population structure

The total number of markers initially included in the association test performed on GEMMA was ~ 1,295,518. From those, 970,131 were included in the association analysis, after filtering for MAF and missingness. The most significant markers were found on chromosome 11 (bp: 51,156,324; p-value: 7.37E-06) and chromosome 27 (bp: 33,027,045; p-value: 7.19E -06) (table 25; figure 15).

Haplotype Analysis

Chromosome 11:

A total of 499 markers were included on the haplotype association analysis for chromosome 11 (bp: 50,585,723 - 52,591,016); total region ~2 Mbp. A higher bayes factor for the haplotype analysis (bf2) was observed between markers 51,255,930 – 51,421,471, suggesting an association between this region and PSSM2 (figure 16). The region encompasses 5 annotated equine genes: *RPL26, KRBA2, RNF222, NDEL1* and *MYH10*. Also, several other equine genes are present nearby.

Chromosome 27:

A total of 469 markers were included on the haplotype association analysis for chromosome 27 (bp: 32,790,007 – 35,930,826); total region ~2 Mbp. No significant haplotype associations were found between this region on chromosome 27 and PSSM2 (figure 17).

Variant discovery from whole genome next generation sequencing

Variants on chromosome 11 from base pair position 51,104,988 to 52,026,206 were annotated and their prevalence evaluated in cases and controls. A total of 4,663 variants, mostly intronic, were present in this region. A total of 18 annotated equine genes are present in the region (*CTC1, PFAS, RANGRF, SLC25A35, ARHGEF15, ODF4, RPL26, KRBA2, RNF222, INDEL1, MYH10, CCDC42, MFSD6L, PIK3R5, PIK3R6, NTN1 and STX8.* figure 16*)*. Based on biological function, predicted functional effect and relative abundance between cases and controls, 4 variants in the *phosphoribosylformylglycinamidine synthase gene (PFAS)* were identified as possible mutations associated with the phenotype (table 26). Three of the mutations were non-synonymous and present in 2 out of 3 cases and none of the controls and 1 mutation was non-synonymous and present in 2 of the 3 cases and one of the 3 controls. All are missense mutations; a C to T substitution that changes the amino acid from threonine to methionine at position 37 (T37M); a G to A substitution that changes the amino acid

from valine to isoleucine at position 307 (V307I); a C to T substitution that changes the amino acid from threonine to isoleucine at position 593 (T593I); a A to G substitution that changes the amino acid from aspargine to serine at position 647 (N647S).

Non-synonymous mutations were also found on genes *CTC1*, *RANGRF*, *ARHGEF15*, *MFSD6L* and *NTN1*, but their abundance was not different between cases and controls.

Discussion

Results presented in Chapter 4 found no detected variants on chromosome 18 in the region of the suggested GWA hit from the 54,000 SNP genome scan that were associated with PSSM2. Failure to identify putative functional variants on ECA18 may be due to the limitations of the approach used to investigate the region or due to incomplete penetrance of the risk allele(s) on ECA18. Genotyping an increased number of cases and controls for all possible variants on ECA18 would be necessary to find potential risk alleles with incomplete penetrance at this locus. Another potential explanation is that the initial association on ECA18 was a false positive association and failure to identify functional alleles would support this assertion.

It is also possible that the true region(s) of the genome harboring PSSM2 risk alleles were not identified due to low power of the GWA analysis. While this 54K SNP chip technology has been successful in finding several genomic regions responsible for important phenotypes[104], it may not be powerful enough for the study of PSSM2. Quarters Horses are known to have lower LD in comparison with other breeds, due to admixture at breed foundation and the rapid population expansion of the breed. GWA analysis relies on LD between the markers and the causative allele. A denser set of SNP markers would be advantageous since it would decrease the distance between the putative causal mutation and the nearest SNP marker, thus likely increasing the LD between the SNP markers and the causal mutation, therefore increasing the power to find a causative mutation. To overcome this issue in a cost-effective way, imputation can be performed and the density of the markers present on the initial GWA can be increased without the need to further genotype the test population, as long as a reference population that was

genotyped on a denser array is available. Here we took advantage of a dense set of ~1.8 million markers genotyped in >500 horses (including 63 Quarter Horses) to impute from ~54k to a higher density to perform a second GWA for PSSM2.

Imputed genotypes are predictions, and not actual observations. Subsequent analyses, such as testing for phenotypic association should incorporate the uncertainty of these predictions, to avoid spurious results. A general approach for this is the use of posterior probabilities for the imputed genotypes, which are probabilities of the three possible genotypes obtained from imputation. Those without uncertainty due to imputation will be 1.0. Genotype probabilities can either be incorporated directly into the statistical analysis (in lieu of the actual genotype) or can be used to assign a best guess genotype. Typically the former is a more powerful approach. One potential pitfall of genotype imputation is unknown genotypic accuracy, which adds an additional level of uncertainty to the statistical analysis. However, since there were 22 horses that were genotyped on both arrays (54k low and 1.8M high-density), the accuracy of the imputation could be evaluated in these horses. This allowed for removal of markers that did not perform well, leaving high quality imputed markers for downstream analysis.

By performing genotype imputation the marker density increased dramatically from 54,000 to ~ 900,000 (after pruning for MAF and marker missingness) and new GWA hits were identified. The previous hit on chromosome 18 disappeared, likely confirming that the association on ECA18 was a false positive in the previous analysis.

The association analysis performed with the imputed data revealed a significant hit, also suggested by the haplotype analysis, on chromosome 11. The *PFAS* gene encodes an enzyme that catalyzes the forth step of IMP (inosine monophosphate) biosynthesis, which is part of the purine de novo nucleotide cycle (synthesis of purine nucleotides requires a total of 10 enzymatic steps). Purines are fundamental biological molecules, serving as components of DNA and RNA, and are essential for processes like intracellular and extracellular signaling, energy metabolism and as coenzymes. Skeletal muscle is a major site of de novo purine production in mammalians. During intense exercise a breakdown of the adenine nucleotide happens in the working muscle through the deamination of AMP to IMP and NH3 (ammonia). It is possible that a mutation in

the *PFAS* gene could lead to a deficiency in the *de novo* purine nucleotide cycle, decreasing purine synthesis in the muscle which could lead to impaired muscle metabolism.

Sequencing the *PFAS* gene in a larger population of cases and controls is necessary to confirm a possible association between this mutation and PSSM2 horses. In conclusion, the imputed GWA followed by association analysis on GEMMA has several advantages in relation to previous GWA analysis. It dramatically increases the number of markers, it considers all 3 genotype probabilities instead of the best guess genotype, and it accounts for population stratification. For those reasons, chromosome 11 is now the top priority location in the genome to look for PSSM2 and further evaluation of variants in the region, in a larger number of horses is essential.

CHR	ID	BP	p_val
27	AX-103844004	33027045	7.19E-06
11	AX-104426812	51156324	7.37E-06
27	AX-104019388	32984360	1.05E-05
11	AX-104121058	51105100	1.08E-05
11	AX-104026085	52025433	1.13E-05
27	AX-104877095	35714262	1.30E-05
11	AX-103715095	52016078	1.37E-05
11	AX-103082061	51135152	1.48E-05
11	AX-103785085	52013360	1.55E-05
11	AX-103817490	52014372	1.75E-05
11	AX-103653838	52044500	1.79E-05
11	AX-103724276	51323761	2.00E-05
11	AX-104120636	51752529	2.44E-05
11	AX-103088090	51176877	2.65E-05
11	AX-104467824	51166724	2.66E-05
11	AX-103037719	51386640	2.88E-05
11	AX-104112223	51103136	2.91E-05

Table 25. List of most significant markers on chromosome 11, from GEMMA. CHR: chromosome; ID: marker name; BP: base pair position; p_val: p-value for marker association with phenotype

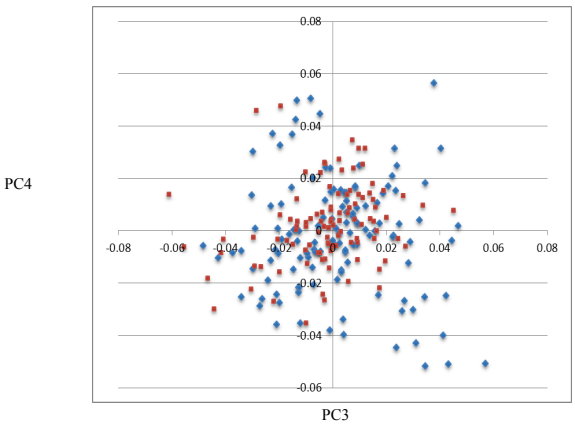


Figure 14. Principal component analysis. Cases (red) and controls (blue). PC3 versus PC4.

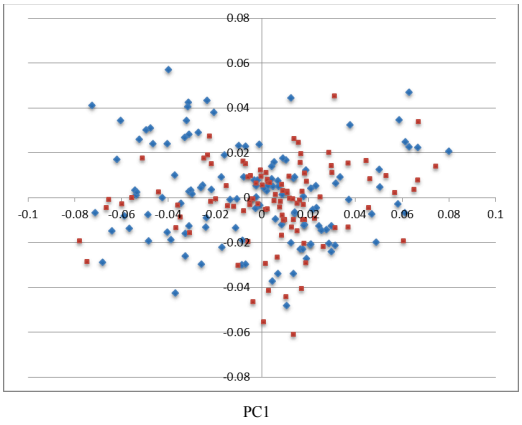


Figure 13. Principal component analysis. Cases (red) and controls (blue). PC1 versus PC3.

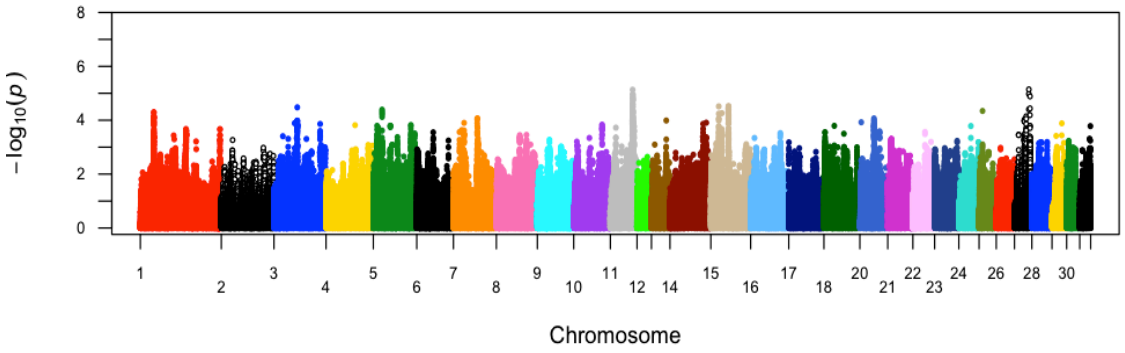


Figure 15. Manhatan plot from imputation and linear mixed model analysis on GEMMA.

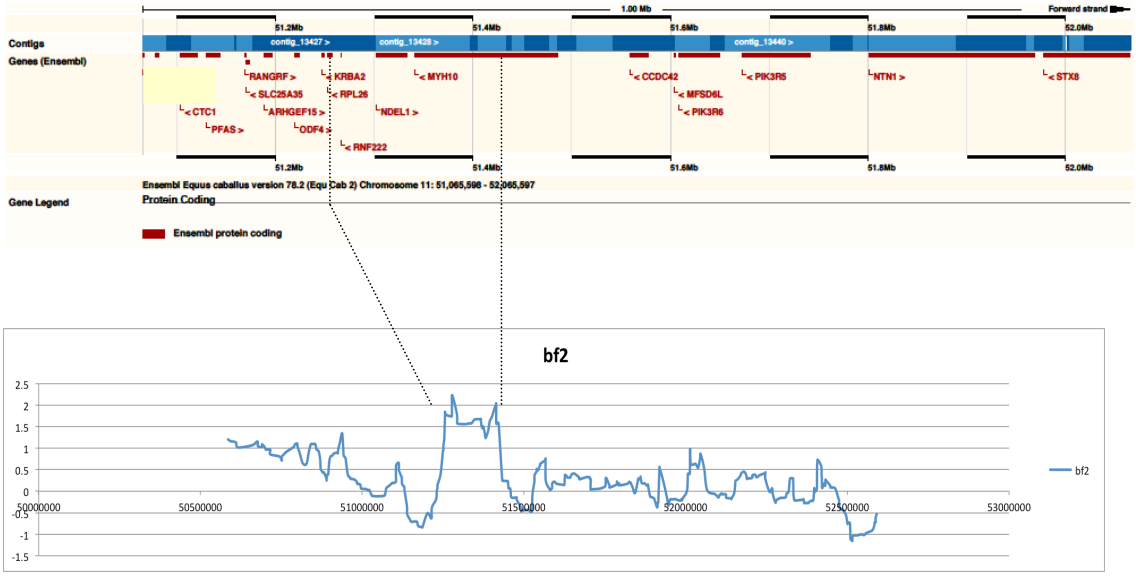


Figure 16. Haplotype analysis on chromosome 11. The $\log_{10}$ Bayes factor (y axis) are plotted at each marker (x axis). The peak area represents the region with the highest haplotype significance. The top figure shows the genes annotated on ensemble in the area with the highest haplotype significance

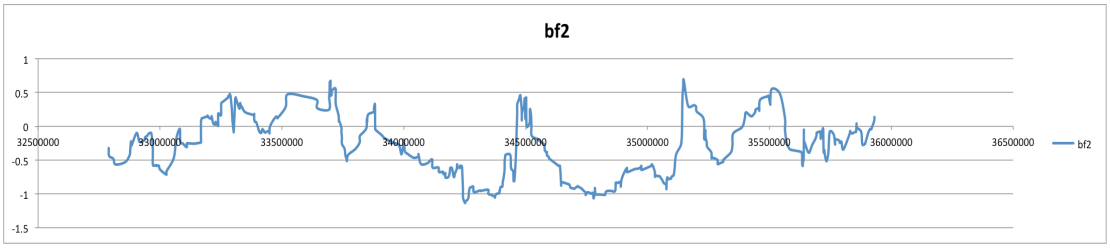


Figure 17. Haplotype analysis on chromosome 27. The log10 Bayes factor(y axis) are plotted at each marker (x axis)

Chr	Position	Mutation	Gene	CT hom	CT het	CT wt	CA hom	CA het	CA wt	p-val
11	51131050	NON_SYNONYMOUS_CODING	*PFAS*	0	0	3	0	2	1	GENO 6.802e-02
11	51133614	NON_SYNONYMOUS_CODING	*PFAS*	0	0	3	0	2	1	GENO 4.163e-02
11	51139603	NON_SYNONYMOUS_CODING	*PFAS*	1	0	2	0	2	1	DOM 3.000e-01
11	51140158	NON_SYNONYMOUS_CODING	*PFAS*	0	0	3	0	2	1	GENO 6.802e-02

Table 26. Significant markers on chromosome 11 and their base pair position and associated gene name. CT hom: control homozygous; CT het: control heterozygous; CT wt: control wild-type; CA hom: case homozygous; CA het: case heterozygous; CA wt: case wild-type; p-val: p-value for variant prevalence between cases and controls.

Chapter 6

Final Considerations and Future Directions

Type 1 Polysaccharide Storage Myopathy

The gene expression profile and enrichment analysis in PSSM1 cases generated hypotheses to the metabolic pathways that are abnormal in those horses, which includes mitochondria biogenesis, oxidative phosphorylation, fatty acid metabolism, glycogen and glucose metabolism (Chapter 3). The study of enrichment analysis is of great value to understand changes in important pathways associated with diseases or conditions. While the results of the enrichment analysis observed in this study were very consistent between experiments, a better understanding of the specific alterations that are occurring in gene regulation in these pathways is necessary. Manual annotation of the pathways based on the disease of interest might be a good solution. Better pathway analysis results would probably be generated if poor annotation were not present on both ends (equine genome and the publicly available pathways). In the case of the equine reference genome, hundreds of transcripts could not be identified despite the effort of blasting them to other species. Some of the available pathways are missing key elements in the physiology of the processes that they describe. To overcome this issue, manually annotating pathways based on what is known about PSSM1 and the genes that are present on the gene expression profile as significant might be of great value. This process could be done by consulting available publications (text mining programs), protein and gene interaction databases.

For a complete description of the equine muscle transcriptome de novo assembly of the raw RNA reads can help identify new transcripts and isoforms, since this approach does not depend in the reference genome. Oases[105] is a de novo transcriptome assembler designed to produce transcripts from for short reads. Oases uploads a preliminary assembly produced by Velvet and clusters the contigs into small groups, called loci. Oases then processes the transcripts and constructs transcript isoforms. For good accuracy of de novo transcripts, high quality data from sequencing are necessary. One important step prior to performing the de novo assembly is to trim reads with low sequencing quality score. Oases runs the de novo assembly in several different k-mer

sizes. At the end (using the merge function) Oases merges transcripts from different k-mer sizes, in an attempt to create longer reads. Once de novo transcripts are identified they can be annotated by blasting to available databases. The new transcripts and isoforms can also be added to the differentially expression analysis to better understand the changes in gene expression in cases and controls. First a new equine transcriptome containing the new transcripts/isoforms is generated and the raw reads are mapped to this reference for expression quantification with the caveat that reads will very frequently not map uniquely to the de novo reference transcriptome. For this reason the use of programs that deal with the mapping uncertainty is necessary to accurately identify changes in gene expression. Corset[106] is a program that performs a gene level analysis of de novo assembled transcripts. It clusters the contigs into "genes" not just based on similarity. It takes in consideration the difference in abundance of raw reads that are present in cases and controls. Corset takes the issue of multi-mapping and uses it as an advantage and produces an output (gene counts) that can be used by Edge R (differentially expressed analysis).

Also the transcriptome profile in this study was measured only at the gene level. When a better equine reference genome becomes available, expression profiles of the genes at the isoform (alternative splicing) levels can be performed. This is the ideal way to measure differences in expression between groups since the same gene can have isoforms that are upregulated and isoforms that are down regulated, causing the final expression for that specific gene to be erroneously unchanged. Creating libraries with longer (100 to 150 bp) and paired end reads would be ideal to detect splice variants.

The need for a better assembled/annotated equine reference genome has become even more apparent with the studies performed in this thesis. To better annotate the reference genome that is already available, AUGUSTUS[107] could be used. AUGUSTUS is a program for gene discovery that incorporates multiple hints into gene prediction models (sources of extrinsic and intrinsic hints). Intrinsic hints include the genomic sequence data and external hints include ESTs (downloaded from available databases) and data generated from RNA-Seq (BAM files). The program was developed into a

pipeline where high-throughput RNA-Seq data was incorporated. This information is used to improve annotation and identification of novel genes.

It is also possible that PSSM2 is not the result of mutation(s) in a protein-encoding gene and is instead a result of alterations in regulatory elements that control gene expression. To be able to look at the gene expression from all different angles, creating a new library that does not restrict the RNA collection to polyA (mRNA) would be ideal to identify possible regulatory elements that might drive changes in gene expression, like small (e.g. miRNA) and long-non coding RNAs (e.g. lincRNAS).

In addition, the study of proteomics and metabolomics, in addition to gene expression, would give the very last and most important piece of information to define the metabolic changes in PSSM1 horses, by looking at potential differences in profile between groups. Proteomics determines the protein content within a wide size range and metabolomics allows a non-target survey of small size metabolites in a biological sample (small-molecules called chemical fingerprints). For example, fatty acids, sugars and protein metabolites can be analyzed and their abundancy can be compared between groups, as well as proteins involved in oxidative metabolism (e.g. NADH ubiquinone), glycolysis, glycogen metabolism (e.g. glycogen phosphorylase) and anaerobic metabolism. In conclusion, the end products of cellular processes can be identified revealing a snap-shot of the physiology of the cell studied, which would be of great importance when comparing cases to controls.

Type 2 Polysaccharide Storage Myopathy

The genetic basis of type 2 PSSM is an area of ongoing research. A previous GWAS using the equine SNP50 array showed a significant marker associated with PSSM2 on equine chromosome 18. Whole genome and target (around chromosome 18) next generation sequencing of cases and controls and BAC clones to fill gaps in the reference genome did not identify mutations in the region that could explain PSSM2 (Chapter 4). To overcome this issue imputation from 50K to 1.8 million markers was

performed and an association study of the new markers revealed a region on chromosome 11 significantly associated with PSSM2 (Chapter 5). This region remained significant after haplotype analysis.

Four non-synonymous mutations of potential interest were identified in the *phosphoribosylformylglycinamidine synthase gene (PFAS).* Three of the mutations were non-synonymous and present in 2 out of 3 cases and none of the controls, and 1 mutation was non-synonymous and present in 2 of the 3 cases and one of the 3 controls. The PFAS gene encodes an enzyme that catalyzes the forth step of IMP (inosine monophosphate) biosynthesis, which is part of the purine de novo nucleotide cycle. Purines are fundamental biological molecules, serving as components of DNA and RNA, and are essential for processes like intracellular and extracellular signaling, energy metabolism and as coenzymes. The equine *PFAS* gene has 1 annotated transcript with 4,017 base pairs and 28 exons. To further evaluate these *PFAS* mutations as a possible cause of PSSM2, primers to amplify and sequence the region encompassing this gene in all PSSM2 cases (104 horses) and controls (124 horses) will be designed. However, as described more below, if PSSM2 is caused by a gene with incomplete penetrance or low-to-moderate relative risk, or if more than one form of PSSM2 exists in the study population, a perfect correlation between the mutation and the phenotype will not be expected.

If none of the four mutations in the *PFAS* gene have significant allele differences between cases and controls a new approach to investigate the region on chromosome 11 will be necessary. A Sequenom assay could be designed to further investigate chromosome 11, including the non-synonymous mutations that were found on genes *CTC1*, *RANGRF*, *ARHGEF15*, *MFSD6L* and *NTN1*. Their abundance was not significantly different between cases and controls on the analysis performed on chapter 5, but sequence data from only 3 horses were present on each group to detect and estimate allele frequency. A Sequenom assay containing all cases (104 horses) and controls (124 horses) would be far more informative. The SNP list can be submitted to GeneSeek to prepare multiplex SNP genotyping assays for the Sequenom platform. The data generated

by Sequenom can then be pruned for genotyping rate and quality and an association analysis can then be performed.

Also, since imputation to the high-density 1.8 million SNP markers is being done in other breeds in our laboratory, a better understanding of which filter parameters should be used to keep just high quality imputed markers in the downstream analysis will be developed. By better filtering imputed markers for accuracy, a chance of finding false positive or false negative markers in the association analysis would decrease. If those approaches still do not answer what causes PSSM2 by finding a highly significant locus with alleles that are highly associated with the disease, increasing the number of cases and controls on the dataset and directly genotyping them on the 670K array might be necessary. This approach would increase the power of the study by increasing the number of cases and controls, and would also increase the marker density without having the genotype uncertainty that is due to the imputation process. Also increasing the number of horses in the study would be needed to find alleles with low-to-moderate risk. The approach described in this paragraph would be of value if the markers on chromosome 11 are false positives and true loci associated with PSSM2 were not detected by the current analysis.

It is also possible that PSSM2 is a polygenic disease or that other forms of excessive glycogen storage diseases exist and the horses called PSSM2 actually represent more than one disease. In both cases the power of the GWAS would be significantly affected. To overcome the issue of a polygenic disease, more cases and controls are necessary to increase the power of the GWAS. To overcome the issue of genetic and phenotypic heterogeneity, a different methodology is necessary to better phenotype the horses prior to association studies. Studying gene expression profile in PSSM2 horses and controls and integrating those findings with proteomics and metabolomics analysis might help cluster PSSM2 horses in different phenotype groups.

Even if all the horses in a PSSM2 GWAS dataset have the same genetic basis for their glycogen storage disease, evaluation of the gene expression in the muscle of horses

with PSSM2 and comparing those changes with controls and PSSM1 horses before and after exercise might provide important information. Considering the use of gene expression profiling along with proteomics and metabolomics to explain what the ultimate changes are in the muscle of PSSM2 horses would be of great value. The metabolome represents the collection of all metabolites in a biological tissue and are the end products of cellular processes. Gene expression data and proteomic analyses reveal the set of gene products being produced in the cell and metabolic profiling can give an instantaneous picture of the physiology of the sampled tissue. The integration of transcriptomic, proteomic and metabolomic information would provide a better understanding of the muscle pathophysiology of PSSM2 horses and, when combined with clinical and other diagnostic data, might enable a better phenotype with which to perform a GWAS.

Reference List

1. Firshman AM, Valberg SJ, Bender JB, Finno CJ. Epidemiologic characteristics and management of polysaccharide storage myopathy in Quarter Horses. Am J Vet Res 2003; 64(10): 1319-27.

2. Valberg SJ, Cardinet GH, III, Carlson GP, DiMauro S. Polysaccharide storage myopathy associated with recurrent exertional rhabdomyolysis in horses. Neuromuscul Disord 1992; 2(5-6): 351-9.

3. Carlstrom B. Uber die atiologie und pathogenese der kreuzlahme des pferdes (Hemaglobinaemia paralytica). Scandinav Archiv 1932; 62: 1-62.

4. Firshman AM, Baird JD, Valberg SJ. Prevalences and clinical signs of polysaccharide storage myopathy and shivers in Belgian draft horses. J Am Vet Med Assoc 2005; 227(12): 1958-64.

5. Valentine BA, Habecker PL, Patterson JS, Njaa BL, Shapiro J, Holshuh HJ et al. Incidence of polysaccharide storage myopathy in draft horse-related breeds: a necropsy study of 37 horses and a mule. J Vet Diagn Invest 2001; 13(1): 63-8.

6. Valentine BA, Cooper BJ. Incidence of polysaccharide storage myopathy: necropsy study of 225 horses. Vet Pathol 2005; 42(6): 823-7.

7. McGowan CM, McGowan TW, Patterson-Kane JC. Prevalence of equine polysaccharide storage myopathy and other myopathies in two equine populations in the United Kingdom. Vet J 2009; 180(3): 330-6.

8. McCue ME, Ribeiro WP, Valberg SJ. Prevalence of polysaccharide storage myopathy in horses with neuromuscular disorders. Equine Vet J Suppl 2006;(36): 340-4.

9. McCue ME, Valberg SJ, Lucio M, Mickelson JR. Glycogen synthase 1 (GYS1) mutation in diverse breeds with polysaccharide storage myopathy. J Vet Intern Med 2008; 22(5): 1228-33.

10. Byrne E, Jones S, Valberg SJ, Zimmel DN, Cohen N. Rhabdomylosis in two foals with polysaccharide storage myopathy and concurrent pneumonia. Compendium on Continuing Education for the Practicing Veterinarian 2000; 22: 503-7.

11. Bloom BA, Valentine BA, Gleed RD, Cable CS. Postanaesthetic recumbency in a Belgian filly with polysaccharide storage myopathy. Vet Rec 1999; 144(3): 73-5.

12. McCue M, Armien A, Lucio M, Mickelson J, Valberg S. Comparative skeletal muscle histopathologic and ultrastructural features in two forms of Polysaccharide Storage Myopathy in horses. Vet Pathol 2009.

13. Firshman AM, Valberg SJ, Bender JB, Annandale EJ, Hayden DW. Comparison of histopathologic criteria and skeletal muscle fixation techniques for the diagnosis of polysaccharide storage myopathy in horses. Vet Pathol 2006; 43(3): 257-69.

14. De La Corte FD, Valberg SJ, MacLeay JM, Mickelson JR. Developmental onset of polysaccharide storage myopathy in 4 Quarter Horse foals. J Vet Intern Med 2002; 16(5): 581-7.

15. Naylor RJ, Livesey L, Schumacher J, Henke N, Massey C, Brock KV et al. Allele copy number and underlying pathology are associated with subclinical severity in equine type 1 polysaccharide storage myopathy (PSSM1). PLoS One 2012; 7(7): e42317.

16. Valentine BA, Credille KM, Lavoie JP, Fatone S, Guard C, Cummings JF, Cooper BJ. Severe polysaccharide storage myopathy in Belgian and Percheron draught horses. Equine Vet J 1997; 29(3): 220-5.

17. Annandale EJ, Valberg SJ, Mickelson JR, Seaquist ER. Insulin sensitivity and skeletal muscle glucose transport in horses with equine polysaccharide storage myopathy. Neuromuscul Disord 2004; 14(10): 666-74.

18. Firshman AM, Baird J.D., Valberg SJ. Prevalence of polysaccharide storage myopathy and "shivers" in Belgian draft horses. J.Vet.Intern.Med. in press. 2006.
Ref Type: Abstract

19. McCue ME, Ribeiro WP, Valberg SJ. The prevalence of polysaccharide storage myopathy in Quarter Horses. J Vet Intern Med 2006; 20(3): 710.

20. McCue ME, Valberg SJ. Estimated prevalence of polysaccharide storage myopathy among overtly healthy Quarter Horses in the United States. J Am Vet Med Assoc 2007; 231(5): 746-50.

21. Valberg SJ, MacLeay JM, Billstrom JA, Hower-Moritz MA, Mickelson JR. Skeletal muscle metabolic response to exercise in horses with 'tying-up' due to polysaccharide storage myopathy. Equine Vet J 1999; 31(1): 43-7.

22. Valberg SJ, Townsend D, Mickelson JR. Skeletal muscle glycolytic capacity and phosphofructokinase regulation in horses with polysaccharide storage myopathy. Am J Vet Res 1998; 59(6): 782-5.

23. Annandale EJ, Valberg SJ, Essen Gustavsson B. The effects of submaximal excercise of adenine nucleotide concentrations in skeletal muscle fibers of horses with polysaccharide storage myopathy. Am J Vet Res 2005; 66(5): 839-45.

24. Dranchak P, Leiper F, Valberg S, Piercy R, Carling D, McCue M, Mickelson J. Biochemical and genetic evaluation of the role of AMPK in equine polysaccharide storage myopathy. Am J Vet Res 2007.

25. McCue ME, Valberg SJ, Miller MB, Wade C, DiMauro S, Akman HO, Mickelson JR. Glycogen synthase (GYS1) mutation causes a novel skeletal muscle glycogenosis. Genomics 2008; 91(5): 458-66.

26. Valberg SJ, MacLeay JM, Billstrom JA, Hower-Moritz MA, Mickelson JR. Skeletal muscle metabolic response to exercise in horses with 'tying-up' due to polysaccharide storage myopathy. Equine Vet J 1999; 31(1): 43-7.

27. J.G Salway. Metabolism at a Glance. Third Edition. 2004. Blackwell Publishing.
Ref Type: Serial (Book,Monograph)

28. Rutherford M, Kaese H, Hendrickson J, Wagner ML, McCue ME, Binns M *et al.* Equine recurrent Uveitis association with the major histocompatibility complex. Plant and Animal Genome XVI. Plant and Animal Genome XVI, 586. 2008.
Ref Type: Conference Proceeding

29. Ribeiro WP, Valberg SJ, Pagan JD, Gustavsson BE. The effect of varying dietary starch and fat content on serum creatine kinase activity and substrate availability in equine polysaccharide storage myopathy. J Vet Intern Med 2004; 18(6): 887-94.

30. Borgia LA, Valberg SJ, McCue ME, Pagan JD, ROE CR. Effect of dietary fats with odd or even numbers of carbon atoms on metabolic response and muscle damage with exercise in Quarter HorseΓÇôtype horses with type 1 polysaccharide storage myopathy. American Journal of Veterinary Research 2010; 71(3): 326-36.

31. Baird JD, Valberg SJ, Anderson SM, McCue ME, Mickelson JR. Presence of the glycogen synthase 1 (GYS1) mutation causing type 1 polysaccharide storage myopathy in continental European draught horse breeds. Vet Rec 2010; 167(20): 781-4.

32. McCue ME, Valberg SJ, Jackson M, Borgia L, Lucio M, Mickelson JR. Polysaccharide storage myopathy phenotype in quarter horse-related breeds is modified by the presence of an RYR1 mutation. Neuromuscul Disord 2009; 19(1): 37-43.

33. Stanley RL, McCue ME, Valberg SJ, Mickelson JR, Mayhew IG, McGowan CM *et al.* A glycogen synthase 1 mutation associated with equine polysaccharide storage myopatyh and exertional rhabdomyolysis occurs in a variety of UK breeds. Equine Vet.J . 2009.
Ref Type: In Press

34. Herszberg B, McCue ME, Larcher T, Mata X, Vaiman A, Chaffaux S et al. A GYS1 gene mutation is highly associated with polysaccharide storage myopathy in Cob Normand draught horses. Anim Genet 2009; 40(1): 94-6.

35. McCue M, Andersen S, Valberg S, Piercy R, Barakzai S, Binn M et al. Estimated prevalence of the Type 1 Polysaccharide Storage Myopathy mutation in selected North American and European breeds. Animal Genetics 2010.

36. Baird JD, Valberg SJ, Anderson SM, McCue ME, Mickelson JR. Presence of the glycogen synthase 1 (GYS1) mutation causing type 1 polysaccharide storage myopathy in continental European draught horse breeds. Vet Rec 2010; 167(20): 781-4.

37. JOHLIG L, Valberg SJ, Mickelson JR, KLUKOWSKA J, REUSSER HR, Straub R, Gerber V. Epidemiological and genetic study of exertional rhabdomyolysis in a Warmblood horse family in Switzerland. Equine Veterinary Journal 2011; 43(2): 240-5.

38. McCoy AM, Schaefer R, Petersen JL, Morrell PL, Slamka MA, Mickelson JR et al. Evidence of positive selection for a glycogen synthase (GYS1) mutation in domestic horse populations. J Hered 2014; 105(2): 163-72.

39. McCoy AM, Schaefer R, Petersen JL, Morrell PL, Slamka MA, Mickelson JR et al. Evidence of positive selection for a glycogen synthase (GYS1) mutation in domestic horse populations. J Hered 2014; 105(2): 163-72.

40. STEPHANIE J.VALBERG. Muscle anatomy, physiology, and adaptations to exercise and training. The Athletic Horse (Second Edition). Principles and Practice of Equine Sports Medicine. ELSEVIER. 2014; 174-201.

41. Eaton MD. Energetics and Performance. In: Hodgson DR, Rose RJ, eds, The Athletic Horse. Philadelphia: W.B. Saunders Company. 1994; 49-61.

42. Freyssenet D. Energy sensing and regulation of gene expression in skeletal muscle. J Appl Physiol 2007; 102(2): 529-40.

43. Haller RG, Vissing J. Functional Evaluation of Metabolic Myopathies. In: Engel AG, Franzini-Armstrong C, eds, Myology. New York: McGraw-Hill. 2004; 665-90.

44. Brojer JT, Essen Gustavsson B, Valberg SJ. Skeletal muscle pro- an macroglycogen in horses with polysaccharide storage myopathy performing light excercise. Am J Vet Res 2006.

45. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008; 5(7): 621-8.

46. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10(1): 57-63.

47. Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet 2011; 12(10): 671-82.

48. Marguerat S, Bahler J. RNA-seq: from technology to biology. Cell Mol Life Sci 2010; 67(4): 569-79.

49. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 2012; 8(2): e1002375.

50. Valentine BA, Flint TH, Fischer KA. Ubiquitin expression in muscle from horses with polysaccharide storage myopathy. Vet Pathol 2006; 43(3): 270-5.

51. Abecasis GR, Ghosh D, Nichols TE. Linkage disequilibrium: ancient history drives the new genetics. Hum Hered 2005; 59(2): 118-24.

52. Hirschhorn JN, Daly MJ. Genome-Wide Association Studies for Common Diseases and Complex Traits. Nat Rev Genet 2005; 6(2): 95-108.

53. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. Nat Protoc 2011; 6(2): 121-33.

54. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet 2006; 7(10): 781-91.

55. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 2009; 326(5954): 865-7.

56. Leeb T, Vogl C, Zhu B, de Jong PJ, Binns MM, Chowdhary BP et al. A human-horse comparative map based on equine BAC end sequences. Genomics 2006; 87(6): 772-6.

57. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM et al. A high density SNP array for the domestic horse and extant Perissodactyla: utility

for association mapping, genetic diversity, and phylogeny studies. PLoS Genet 2012; 8(1): e1002451.

58. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010; 11(7): 499-511.

59. McCoy AM, McCue ME. Validation of imputation between equine genotyping arrays. Anim Genet 2014; 45(1): 153.

60. Egan B, Zierath JR. Exercise metabolism and the molecular regulation of skeletal muscle adaptation. Cell Metab 2013; 17(2): 162-84.

61. Park KD, Park J, Ko J, Kim BC, Kim HS, Ahn K et al. Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. BMC Genomics 2012; 13: 473.

62. McGivney BA, McGettigan PA, Browne JA, Evans AC, Fonseca RG, Loftus BJ et al. Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. BMC Genomics 2010; 11: 398.

63. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 2008; 18(9): 1509-17.

64. Page GP, Edwards JW, Barnes S, Weindruch R, Allison DB. A design and statistical perspective on microarray gene expression studies in nutrition: the need for playful creativity and scientific hard-mindedness. Nutrition 2003; 19(11-12): 997-1000.

65. Tempelman RJ. Assessing statistical precision, power, and robustness of alternative experimental designs for two color microarray platforms based on mixed effects models. Veterinary Immunology and Immunopathology 2005; 105(3-4): 175-86.

66. Illumina Tru_seq. Illumina website . 2015.
Ref Type: Electronic Citation

67. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009; 25(9): 1105-11.

68. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012; 7(3): 562-78.

69. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012; 7(3): 562-78.

70. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012; 7(3): 562-78.

71. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 2013; 8(9): 1765-86.

72. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 2005; 102(43): 15545-50.

73. Yang Y, Creer A, Jemiolo B, Trappe S. Time course of myogenic and metabolic gene expression in response to acute exercise in human skeletal muscle. J Appl Physiol 2005; 98(5): 1745-52.

74. Parker GE, Pederson BA, Obayashi M, Schroeder JM, Harris RA, Roach PJ. Gene expression profiling of mice with genetically modified muscle glycogen content. Biochem J 2006; 395(1): 137-45.

75. Barrey E, Mucher E, Jeansoule N, Larcher T, Guigand L, Herszberg B et al. Gene expression profiling in equine polysaccharide storage myopathy revealed inflammation, glycogenesis inhibition, hypoxia and mitochondrial dysfunctions. BMC Vet Res 2009; 5: 29.

76. Marguerat S, Bahler J. RNA-seq: from technology to biology. Cell Mol Life Sci 2010; 67(4): 569-79.

77. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012; 7(3): 562-78.

78. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 2013; 8(9): 1765-86.

79. CLC Genomics Workbench 7.0.3 (http://www.clcbio.com). CLC Genomics Workbench 7.0.3 (http://www.clcbio.com) . 2015.
Ref Type: Electronic Citation

80. Sequencher® version 5.3 sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA. http://www.genecodes.com . 2015. Ref Type: Electronic Citation

81. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011; 43(5): 491-8.

82. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20(9): 1297-303.

83. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26(5): 589-95.

84. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25(14): 1754-60.

85. http://broadinstitute.github.io/picard. Broad Institute . 2015. Ref Type: Electronic Citation

86. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20(9): 1297-303.

87. Cingolani P, Platts A, Wang lL, Coon M, Nguyen T, Wang L et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012; 6(2): 80-92.

88. Cingolani P, Platts A, Wang lL, Coon M, Nguyen T, Wang L et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012; 6(2): 80-92.

89. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009; 84(2): 210-23.

90. Khatkar MS, Moser G, Hayes BJ, Raadsma HW. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. BMC Genomics 2012; 13: 538.

91. Browning BL, Browning SR. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. Genet Epidemiol 2007; 31(5): 365-75.

92. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. The variant call format and VCFtools. Bioinformatics 2011; 27(15): 2156-8.

93. Browning BL, Browning SR. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. Genet Epidemiol 2007; 31(5): 365-75.

94. Browning BL, Browning SR. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. Genet Epidemiol 2007; 31(5): 365-75.

95. Browning BL, Browning SR. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. Genet Epidemiol 2007; 31(5): 365-75.

96. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012; 44(7): 821-4.

97. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012; 44(7): 821-4.

98. Xu H, Guan Y. Detecting local haplotype sharing and haplotype association. Genetics 2014; 197(3): 823-38.

99. Xu H, Guan Y. Detecting local haplotype sharing and haplotype association. Genetics 2014; 197(3): 823-38.

100. Xu H, Guan Y. Detecting local haplotype sharing and haplotype association. Genetics 2014; 197(3): 823-38.

101. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012; 44(7): 821-4.

102. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81(3): 559-75.

103. LI H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25(14): 1754-60.

104. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J et al. Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS Genet 2013; 9(1): e1003211.

105. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 2012; 28(8): 1086-92.

106. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biol 2014; 15(7): 410.

107. Li L, Chen E, Yang C, Zhu J, Jayaraman P, De PJ et al. Improved rat genome gene prediction by integration of ESTs with RNA-Seq information. Bioinformatics 2015; 31(1): 25-32.

108. Borgia L, Valberg S, McCue M, Watts K, Pagan J. Glycaemic and insulinaemic responses to feeding hay with different non-structural carbohydrate content in control and polysaccharide storage myopathy-affected horses. J Anim Physiol Anim Nutr (Berl) 2011; 95(6): 798-807.

}